PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	FOR FURTHER ACTION see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
JMH/397/PCT		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/GB 00/02286	12/06/2000	11/06/1999
Applicant		
QXYZ LIMITED		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 8 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language,** the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

   [ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

   b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

   [ ] contained in the international application in written form.

   [ ] filed together with the international application in computer readable form.

   [X] furnished subsequently to this Authority in written form.

   [X] furnished subsequently to this Authority in computer readble form.

   [X] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   [X] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [ ] **Certain claims were found unsearchable** (See Box I).

3. [X] **Unity of invention is lacking** (see Box II).

4. With regard to the **title,**

   [ ] the text is approved as submitted by the applicant.

   [X] the text has been established by this Authority to read as follows:

   DNA MANIPULATION METHODS, APPLICATIONS FOR SYNTHETIC ENZYMES AND USE FOR POLYKETIDE PRODUCTION.

5. With regard to the **abstract,**

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. ______

   [ ] as suggested by the applicant.

   [ ] because the applicant failed to suggest a figure.

   [ ] because this figure better characterizes the invention.

   [X] None of the figures.

Box I Observati ns where certain claims wer found unsearchabl (Continuation of it m 1 of first she t)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: (1-8,32,33,37,38)-complete, (16-31,36,41-43, 46-48)-partially

A method of assembling several DNA units in sequence in a DNA construct, which method comprises the step of: a) providing each DNA unit with a restriction enzyme recognition sequence at its 5' end and with a recognition sequence for the same restriction enzyme at its 3' end that is combined with a restriction site for a DNA modification enzyme, b) providing a starting DNA construct having an accessible restriction site for the same or a compatible restriction enzyme and cleaving the starting DNA construct with a restriction enzyme, c) inserting the desired DNA unit and bringing the ligated product into contact with a DNA modification enzyme such that the restriction site at the 3' end of the inserted DNA unit is abolished, d) cleaving the ligated product at an accessible unmodified recognition site for the same or a compatible restriction enzyme, e) repeating step c) and d) to introduce each desired DNA unit to give a DNA construct containing all the desired units in sequence; DNA construct incorporating one or more DNA assemblies encoding synthetic enzymes and/or hosts expressing DNA constructs made by said method; compounds produced by synthetic enzymes encoded by said DNA assemblies; a method of synthesising a target molecule using said method; a method of making a synthetic enzyme to catalyse the synthesis of a target molecule using said method; a library of DNA units encoding a catalytic or transport protein domains, wherein each DNA unit has a recognition sequence for a restriction enzyme at its 5'-end and a second recognition sequence for the same or a compatible enzyme at its 3'-end which incorporates a recognition sequence for a DNA modifying enzyme;
a module comprising a DNA sequence encoding a functional set of polyketide synthetic domains wherein the module has a recognition sequence for a restriction enzyme at its 5'-end and a second recognition sequence for the same or a compatible enzyme at its 3'-end which incorporates a recognition sequence for a DNA modifying enzyme;
a method of transforming a host with one or more synthetic DNA assemblies encoding enzyme domains, wherein the DNA assemblies are said modules;

2. Claims: (9-15,33,34,39,40)-complete, (16-31,36,41-43, 46-48)-partially

Idem as invention 1, but limited to a method of: assembling several DNA units in sequence in a DNA construct, which method comprises the step of: a) providing a first DNA unit

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

with a recognition sequence for a first restriction enzyme at its 3' end, and cleaving the said first DNA unit with said first restriction enzyme, b) providing each other DNA unit with a recognition sequence at its 5' end for a second restriction enzyme which has a compatible ligation sequence with that of the first restriction enzyme, and a downstream recognition sequence for said first restriction enzyme followed by a downstream recognition sequence for a third restriction enzyme at its 3' end, and cleaving each said other DNA unit with the second and third restriction enzymes, c) ligating the said first DNA unit with a desired other DNA unit to form a ligated product such that the ligation of the two units abolishes the recognition site for the first restriction enzyme at the ligation junction, and cleaving the ligated product with said first restriction enzymel, d) ligating the product from c) with a desired DNA unit from b) to form a ligated product and cleaving the ligated product with said first restriction enzyme, e)repeating step d) with each other DNA unit in turn so as to assemble the DNA unit in sequence;

3. Claims: 44-45

A method of transforming a host with one or more synthetic DNA assemblies encoding enzyme domains which comprises the step of: a) Inserting said DNA assembly into a vector containing a mutated internal fragment of a recA gene sequence such that the vector is capable of undergoing homologous recombination with the recA gene of the host, b) bringing said vector into contact with a host chromosome under conditions which permit homologous recombination to take place, c) disrupting the host recA gene by integration of the DNA of said vector into the chromosome; said method wherein the expression vector is used to transform a Streptomyces host;

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/02286

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12N15/66 C12N15/52 C12N15/90 C12P17/06
C12P17/08

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, PAJ, CAB Data, STRAND, EPO-Internal, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 17811 A (CHROMAXOME CORP) 30 April 1998 (1998-04-30) claims 1-26; figure 5E; example 5.5.5.	22-26
X	ROWE C J ET AL: "Construction of new vectors for high-level expression in actinomycetes" GENE,NL,ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, vol. 216, no. 1, August 1998 (1998-08), pages 215-223, XP004149299 ISSN: 0378-1119 cited in the application the whole document	22-26

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 February 2001	27.02.01

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Hornig, H

3

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 49315 A (KOSAN BIOSCIENCES INC ;UNIV LELAND STANFORD JUNIOR (US)) 5 November 1998 (1998-11-05) claims 1-24; figure 6A	22-26
X	WO 96 40968 A (UNIV LELAND STANFORD JUNIOR ;JOHN INNES CENTRE (GB)) 19 December 1996 (1996-12-19) the whole document	22-26
X	MCDANIEL R ET AL: "Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel unnatural natural products" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA,NATIONAL ACADEMY OF SCIENCE. WASHINGTON,US, vol. 96, no. 5, March 1999 (1999-03), pages 1846-1851, XP002143433 ISSN: 0027-8424 the whole document	22-26
X	MUTH G ET AL: "Mutational analysis of Streptomyces lividans recA gene suggests that only mutants with residual activity remain viable." MOLECULAR & GENERAL GENETICS, vol. 255, no. 4, 1997, pages 420-428, XP002160032 ISSN: 0026-8925 the whole document	44,45
X	EP 0 841 402 A (NAT INST AGROBIO RES) 13 May 1998 (1998-05-13) page 3, line 28 - line 33; claim 19	44
X	US 4 963 487 A (SCHIMMEL PAUL R) 16 October 1990 (1990-10-16) column 2, line 9 - line 55	44
X	US 4 713 337 A (JASIN MARIA ET AL) 15 December 1987 (1987-12-15) column 4, line 66 -column 6, line 5	44
A	US 5 863 730 A (MASSON ET AL.) 26 January 1999 (1999-01-26) claims 1-17	
T	WO 00 63360 A (CELLTECH THERAPEUTICS LTD ;FINNEY HELENE MARGARET (GB); LAWSON ALA) 26 October 2000 (2000-10-26) the whole document	

-/--

3

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 98 38326 A (ZINK MARY ANN ;XU GUOPING (US); HODGSON CLAGUE P (US); NATURE TECH) 3 September 1998 (1998-09-03) the whole document ---	
A	WO 97 28282 A (STRATAGENE INC) 7 August 1997 (1997-08-07) the whole document ---	
A	TER HAAR ERNST ET AL: "Discodermolide, a cytotoxic marine agent that stabilizes microtubules more potently than taxol." BIOCHEMISTRY, vol. 35, no. 1, 1996, pages 243-250, XP002154629 ISSN: 0006-2960 cited in the application the whole document ---	
A	NERENBERG J B ET AL: "TOTAL SYNTHESIS OF THE IMMUNOSUPPRESSIVE AGENT (-)-DISCODERMOLIDE" JOURNAL OF THE AMERICAN CHEMICAL SOCIETY,US,AMERICAN CHEMICAL SOCIETY, WASHINGTON, DC, vol. 115, no. 26, 1993, pages 12621-12622, XP000652058 ISSN: 0002-7863 the whole document ---	
A	TAPIOLAS D M ET AL: "OCTALACTINS A AND B CYTOTOXIC EIGHT-MEMBERED-RING LACTONES FROM A MARINE BACTERIUM STREPTOMYCES-SP" JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 113, no. 12, 1991, pages 4682-4683, XP002154630 ISSN: 0002-7863 cited in the application the whole document ---	
A	YAMADA SHINYA ET AL: "Total synthesis of (-)-decarestrictine J." BIOSCIENCE BIOTECHNOLOGY AND BIOCHEMISTRY, vol. 59, no. 9, 1995, pages 1657-1660, XP002154631 ISSN: 0916-8451 cited in the application the whole document ---	

-/--

3

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,A	RANGANATHAN ANAND ET AL: "Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: A route to simple statin analogues." CHEMISTRY & BIOLOGY (LONDON), vol. 6, no. 10, October 1999 (1999-10), pages 731-741, XP000971117 ISSN: 1074-5521 the whole document -----	

3

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 00/02286

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
WO 9817811	A	30-04-1998	US	5783431	A	21-07-1998
			AU	5163298	A	15-05-1998
			EP	0951557	A	27-10-1999
WO 9849315	A	05-11-1998	AU	7172298	A	24-11-1998
			EP	0979286	A	16-02-2000
			US	6117659	A	12-09-2000
WO 9640968	A	19-12-1996	US	5712146	A	27-01-1998
			AU	703920	B	01-04-1999
			AU	6157596	A	30-12-1996
			CA	2224104	A	19-12-1996
			EP	0871760	A	21-10-1998
			NZ	310729	A	29-09-1999
			US	6077696	A	20-06-2000
			US	5962290	A	05-10-1999
EP 0841402	A	13-05-1998	AU	694393	B	16-07-1998
			AU	3922697	A	02-04-1998
			AU	721577	B	06-07-2000
			AU	6804898	A	30-07-1998
			CA	2216596	A	26-03-1998
			CN	1182796	A	27-05-1998
			JP	10155485	A	16-06-1998
			US	6165780	A	26-12-2000
US 4963487	A	16-10-1990	US	4713337	A	15-12-1987
			US	4774180	A	27-09-1988
			EP	0257095	A	02-03-1988
			JP	63502723	T	13-10-1988
			WO	8705331	A	11-09-1987
US 4713337	A	15-12-1987	US	4963487	A	16-10-1990
US 5863730	A	26-01-1999	FR	2738841	A	21-03-1997
			EP	0771872	A	07-05-1997
WO 0063360	A	26-10-2000	NONE			
WO 9838326	A	03-09-1998	AU	6443298	A	18-09-1998
WO 9728282	A	07-08-1997	NONE			

PATENT COOPERATION TREATY

From the
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

by fax and post

PCT

To:

HEATON, Joanne M.
STEVENS, HEWLETT & PERKINS
1 St. Augustine's Place
Bristol BS1 4UD
GRANDE BRETAGNE

STEVENS HEWLETT
& PERKINS
BRISTOL

10 OCT 2001

DIA
FILE
No.

FAX: 0044 117 922 6004

NOTIFICATION OF TRANSMITTAL OF
THE INTERNATIONAL PRELIMINARY
EXAMINATION REPORT

(PCT Rule 71.1)

Date of mailing (*day/month/year*) 05.10.2001

Applicant's or agent's file reference	
JMH/397/PCT	**IMPORTANT NOTIFICATION**

International application No.	International filing date (*day/month/year*)	Priority date (*day/month/year*)
PCT/GB00/02286	12/06/2000	11/06/1999

Applicant
QXYZ LIMITED et al.

1. The applicant is hereby notified that this International Preliminary Examining Authority transmits herewith the international preliminary examination report and its annexes, if any, established on the international application.

2. A copy of the report and its annexes, if any, is being transmitted to the International Bureau for communication to all the elected Offices.

3. Where required by any of the elected Offices, the International Bureau will prepare an English translation of the report (but not of any annexes) and will transmit such translation to those Offices.

4. **REMINDER**

   The applicant must enter the national phase before each elected Office by performing certain acts (filing translations and paying national fees) within 30 months from the priority date (or later in some Offices) (Article 39(1)) (see also the reminder sent by the International Bureau with Form PCT/IB/301).

   Where a translation of the international application must be furnished to an elected Office, that translation must contain a translation of any annexes to the international preliminary examination report. It is the applicant's responsibility to prepare and furnish such translation directly to each elected Office concerned.

   For further details on the applicable time limits and requirements of the elected Offices, see Volume II of the PCT Applicant's Guide.

Name and mailing address of the IPEA/	Authorized officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Guerin, A Tel.+49 89 2399-8061

PATENT COOPERATION EATY

PCT

REC'D 10 OCT 2001

WIPO

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference JMH/397/PCT	**FOR FURTHER ACTION**	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/GB00/02286	International filing date *(day/month/year)* 12/06/2000	Priority date *(day/month/year)* 11/06/1999
International Patent Classification (IPC) or national classification and IPC C12N9/10		
Applicant QXYZ LIMITED et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 7 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 34 sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☒ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☒ Certain observations on the international application

Date of submission of the demand 10/01/2001	Date of completion of this report 05.10.2001
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Ury, A Telephone No. +49 89 2399 8411

I. Basis of the report

1. With regard to the **elements** of the international application *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17))*:

Description, pages:

1-21a	as originally filed			
22-45	as received on	01/11/2000	with letter of	18/10/2000

Claims, No.:

1-48	as received on	01/11/2000	with letter of	18/10/2000

Drawings, sheets:

1/23-23/23	as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☐ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

   ☐ the description, pages:
   ☐ the claims, Nos.:
   ☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

   (Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

IV. Lack of unity of invention

1. In response to the invitation to restrict or pay additional fees the applicant has:

   ☐ restricted the claims.

   ☒ paid additional fees.

   ☐ paid additional fees under protest.

   ☐ neither restricted nor paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied and chose, according to Rule 68.1, not to invite the applicant to restrict or pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rules 13.1, 13.2 and 13.3 is

   ☐ complied with.

   ☐ not complied with for the following reasons:

4. Consequently, the following parts of the international application were the subject of international preliminary examination in establishing this report:

   ☒ all parts.

   ☐ the parts relating to claims Nos. .

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-21, 27, 28, 31-43, 46, 47
	No:	Claims	22-26, 29, 30, 44, 45, 48
Inventive step (IS)	Yes:	Claims	1-21, 27, 28, 31
	No:	Claims	22-26, 29, 30, 32-48
Industrial applicability (IA)	Yes:	Claims	1-48
	No:	Claims	

2. *Citations and explanations*
   see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the *claims are fully supported by the description, are made:*
see separate sheet

Item V.

Reference is made to the following documents:

D1: WO 98 17811 A (CHROMAXOME CORP) 30 April 1998 (1998-04-30)
D2: ROWE C J ET AL: 'Construction of new vectors for high-level expression in actinomycetes' GENE,NL,ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, vol. 216, no. 1, August 1998 (1998-08), pages 215-223, XP004149299 ISSN: 0378-1119 cited in the application
D3: WO 98 49315 A (KOSAN BIOSCIENCES INC ;UNIV LELAND STANFORD JUNIOR (US)) 5 November 1998 (1998-11-05)
D4: WO 96 40968 A (UNIV LELAND STANFORD JUNIOR ;JOHN INNES CENTRE (GB)) 19 December 1996 (1996-12-19)
D5: MCDANIEL R ET AL: 'Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel unnatural natural products' PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA,NATIONAL ACADEMY OF SCIENCE. WASHINGTON,US, vol. 96, no. 5, March 1999 (1999-03), pages 1846-1851, XP002143433 ISSN: 0027-8424
D6: MUTH G ET AL: 'Mutational analysis of Streptomyces lividans recA gene suggests that only mutants with residual activity remain viable.' MOLECULAR & GENERAL GENETICS, vol. 255, no. 4, 1997, pages 420-428, XP002160032 ISSN: 0026-8925
D7: EP-A-0 841 402 (NAT INST AGROBIO RES) 13 May 1998 (1998-05-13)
D8: US-A-4 963 487 (SCHIMMEL PAUL R) 16 October 1990 (1990-10-16)
D9: US-A-4 713 337 (JASIN MARIA ET AL) 15 December 1987 (1987-12-15)

I) The methods of assembling several DNA units according to present claims 1-21 appear to be novel and to involve an inventive step (Article 33.2 and 3 PCT) in view of the prior art documents cited in the International Search Report.
As well, the methods according to claims 27, 28 and 31 also fulfil the requirements of Article 33.2 and 3 PCT since they involve the use of any one of the methods of claims 1-21.

II) However, the products according to claims 22-26, obtained by the methods of claims 1-21 are not different from the same products disclosed in the prior art, although said prior art products have been obtained by other methods.
Thus, the disclosures of D1 (see the claims, fig. 5E and example 5.5.5), D2 (see abstract), D3, D4 and D5 (see abstract) are novelty destroying for the subject-matter of present claims 22-26 (Article 33.2 PCT).

The same reasoning applies *mutatis mutandis* for the subject-matter of present claims 29 and 30 which also lack novelty over D1-D5 (Article 33.2 PCT).

III) The method according to claim 44 (i.e. integration of expression plasmids into a host using a mutated internal fragment of the recA gene as the region for homologous recombination) was known per se in the art as shown by documents D6-D9 (see the passages mentioned in the ISR). Claim 44 indicates that the host is transformed with "one or more synthetic DNA assemblies encoding enzyme domains". Any DNA sequence encoding an enzyme falls under this definition. Thus, D6-D9 which disclose said method with plasmids containing sequences encoding other enzymes (for instance the rep sequence of D6, see fig. 2A) destroy the novelty of claim 44.
Document D6 discloses said method to transform a *Streptomyces* host. Thus, D6 also anticipates claim 45 (Article 33.2 PCT).

IV) Claim 48 lacks novelty and inventive step under Article 33.2 and 3 PCT, for obvious reasons.

V)1) In the light of the prior art, the problem underlying part of the present application (claims 1-21) can be defined as the provision of further alternative methods of assembling several DNA units in sequence. Two methods (claims 1-8 and claims 9-15) involving an inventive step offer two alternative solutions to this problem. However, the libraries and modules according to claims 32-36 and 37-41, respectively merely consist in intermediate products which may or may not be used in the above mentioned methods. These products per se do not provide any solution to the underlying problem and therefore cannot be considered as involving an inventive step (Article 33.3 PCT).

2) Claims 42, 43, 46, 47 do not contain any features which, in combination with the features of any claim to which they refer, meet the requirements of the PCT in respect of inventive step (Article 33.3 PCT).

Item VIII.

The passage "(see ... paper)" in claim 16 is not allowable under Article 6 PCT.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 December 2000 (21.12.2000)

PCT

(10) International Publication Number
WO 00/77181 A2

(51) International Patent Classification[7]: C12N 9/10, 15/52, 15/10, 15/90, C12P 17/00

(21) International Application Number: PCT/GB00/02286

(22) International Filing Date: 12 June 2000 (12.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9913694.7 11 June 1999 (11.06.1999) GB

11 Dec 01 30 mos

(71) Applicant (for all designated States except US): QXYZ LIMITED [GB/GB]; Mitchells Roberton, Solicitors, George House, 36 North Hanover Street, Glasgow G1 2AD (GB).

(72) Inventor; and
(75) Inventor/Applicant (for US only): RANGANATHAN, Anand [IN/IN]; International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, P.O. Box 10504, New Delhi 110067, Maharashtra (IN).

(74) Agent: HEATON, Joanne, Marie; Stevens, Hewlett & Perkins, 1 St Augustine's Place, Bristol BS1 4UD (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/77181 A2

(54) Title: DNA MANIPULATION METHODS AND APPLICATIONS FOR SYNTHETIC ENZYMES

(57) Abstract: The invention comprises a method of assembling several DNA units in sequence in a DNA construct and all derivatives of this method. In particular the production of synthetic enzymes is contemplated. Each DNA unit is provided with the same restriction enzyme recognition site at its 5' and 3' ends. The restriction recognition site at its 3' end being combined with a recognition site for a DNA modification enzyme. A DNA construct having the same or a compatible accessible restriction site, as provided in the DNA unit, is cleaved at the restriction site by the appropriate restriction enzyme. The desired DNA unit is then inserted into the DNA construct, this ligated product subsequently being brought into contact with a DNA modification enzyme such that the restriction site at the 3' end of the inserted DNA unit is abolished. The ligated product is then cleaved at the remaining unmodified restriction recognition site and a subsequent DNA unit is inserted. This process is repeated introducing each desired DNA unit to give a DNA construct containing all the desired units in sequence.

DNA MANIPULATION METHODS AND APPLICATIONS FOR SYNTHETIC ENZYMES.

Background

Polyketides, including the valuable drugs avermectin, erythromycin and rapamycin, are natural products that are synthesised by stepwise condensation of acetate, propionate and occasionally butyrate units. The enzymes that take part in the biosynthesis of polyketide chains are collectively known as the polyketide synthase (PKS). PKSs include examples of both type I (multifunctional enzyme) and type II (dissociable complex) organisation. The sequencing of the gene clusters encoding the erythromycin- (ery) and rapamycin- (rap) producing polyketide synthases has shown that each cycle of polyketide chain extension is catalysed by a different set or 'module' of enzyme activities, housed in a few very large multienzyme polypeptides. The basic building blocks of modules are enzymatic 'domains' that are covalently linked together. The ability of these domains to act upon the carbon chain and remove/add functionalities is reminiscent of a molecule being acted upon by chemical reagents in a chemical synthesis. The aim is therefore to assemble these domains or even modules in a manner as desired, so that the linked enzymes can carry out efficient synthesis of any target molecule. Until now, it has however not been possible to find a versatile methodology to assemble these PKS units.

The whole area of polyketide research is at a stage where the flexibility of the whole enzymatic machinery is understood, despite the lack of any X-ray crystal structure data on these giant enzymes, but it remains difficult to "re-assemble" the enzymes *de novo*. A *de novo* synthesis is desirable for two reasons. Firstly, one does not need to change the structure of, for example, an antibiotic using tedious chemical methodologies that are time-consuming and expensive. Engineering an

synthetic enzyme at the genetic level is much easier, faster and cheaper. As more and more antibiotics are rendered useless, simply because the bacteria they were active against have developed ways in which to become resistant to these drugs, there is an urgency to keep developing altered drug structures. Secondly, there is an ever-growing need for new drugs, more potent in their action than their predecessors. Whilst nature provides a large proportion of the new molecules that are, for example, antibiotic, anticholesterol, antifungal, or anti-cancer, the complicated structures of these drugs (for example the anti-cancer Taxol) makes it increasingly difficult for chemists to carry out conventional syntheses. The problem is made more difficult by the fact that the genes that make these drugs cannot always be isolated.

The isolation of the genes coding for the proteins that make the highly potent anti-cancer compound Taxol, has not as yet been reported. The resulting choice for obtaining Taxol is either to cut down 200 Pacific Yew trees to obtain enough taxol for one chemotherapy session, or to make the drug chemically using one of the many exceedingly expensive and long chemical routes that have appeared recently in the literature.

With the isolation, cloning and sequencing of the genes coding for the erythromycin polyketide synthases, a model for the functioning of modular type I PKSs began to emerge. It was clear that such a system is genetically programmed to carry out the necessary catalytic activities needed for processing of the polyketide chain. It is hypothesised that each domain acts independently on the progressing carbon skeleton and there is a correlation between the structure of the growing chain and the enzymatic activities carried out by the enzymes.

The first conclusive proof of such an arrangement came from experiments done by Donadio *et al.* (1991, 1993). One such experiment (1991) involved an in-frame deletion in the ORF3 segment of erythromycin chromosome. This deletion eliminated the entire 183 amino acids of the ketoreductase domain of *ery* PKS module 5, along with some of the

flanking region (a total of 271 amino acids) and resulted in the production of 5,6-dideoxy-3-α-mycarosyl-5-oxo-erythronolide B, the structure of which was confirmed by X-ray crystallography. Replacement of two amino acids in the putative NAD(P)H-binding motif of the enoylreductase domain encoded by ORF2 resulted in a new macrolide $\Delta^{6,7}$-anhydroerythromycin C being produced albeit in low yield. These results demonstrated that erythromycin PKS can be genetically reprogrammed to produce novel macrolides that would otherwise be difficult to get via chemical means.

During the analysis of the fermentation products produced by a strain of *S. erythraea* that was genetically engineered to produce an analogue of 6dEB, it was found that a minor component of the fermentation was 3,5-dihydroxy-2,4-dimethyl-*n*-heptanoic acid δ-lactone (Donadio *et al.*, 1991). This product was predicted to result from premature release of the chain from either the ACP of module 2 or the KS of module 3. A greater yield of this triketide product was obtained by heterologous over-expression of ORF1 in *Streptomyces coelicolor* (Kao *et al.*, 1994), which also showed that DEBS1 can function autonomously. More recently (Cortés *et al.*, 1995), a six-membered lactone was produced through genetically engineering the PKS. By repositioning the TE (cyclase) domain from module 6 to the C-terminus of module 2 (end of DEBS1), it was found that the yield of the lactone is increased by five-fold to 10-15 mg/L as compared to 1-3 mg/L obtained by Kao *et al.*

The relocation of the thioesterase domain at the end of DEBS1 was the first example demonstrating the efficacy of repositioning domains in type I modular systems. Since then, numerous such experiments have been carried out in order to probe further the efficacy of these multienzymes. The TE domain has been relocated at the end of module 5 as well as module 3 respectively (Kao *et al.*, 1995, 1996). In both cases, the predicted compounds were produced that resulted from truncation of the progressing polyketide chain. Release of the 12-membered product in the former case showed that the thioesterase domain can indeed catalyse

ring closure even for less energetically favourable reactions. In the second experiment, two products were produced, one of them thought to be resulting from spontaneous decarboxylation.

The first example of a chimaeric polyketide synthase constructed from a domain taken from a second PKS was demonstrated by Oliynyk *et al.* (1996). An acyltransferase domain (AT) from module 2 of the rapamycin polyketide synthase was used to replace the AT of module 1 in the DEBS1-TE system. The resulting triketide lactone had a methyl group missing at position 5 of the six-membered ring. This was expected since the AT of module 2 of *rap* PKS (unlike the AT of module 1 of DEBS1) incorporates a malonyl-CoA extender unit, instead of a methylmalonyl-CoA unit.

Thus, it has been shown that not only can domains residing within a particular PKS be interchanged or destroyed, analogous domains can be derived from other synthases for the same purpose or for achieving the required synthetic goal. Such a strategy immediately provides a glimpse of the manner in which "designer" polyketides can be constructed through using "off-the-shelf" gene products.

More recently, another hybrid system has been constructed (Marsden *et al.*, 1998) wherein a complete loading module from the avermectin PKS has been swapped with the erythromycin loading module, while keeping the rest of the DEBS modules intact. As expected, incorporation of butyryl-CoA as well as 2-methylisobutyryl-CoA was seen and in both cases, the end products contained the above mentioned residues. A closely-related experiment has been reported by Kuhstoss *et al.* (1996) in which the loading module from the platenolide PKS was replaced with the loading module from tylactone PKS to yield the expected polyketide product.

It is very clear from the various engineering efforts outlined above that the aim must now be to exploit the potential for genetic manipulation of type I (modular) polyketide synthases (PKS) to produce hybrid synthases

that might catalyse the formation of novel secondary metabolites in a predictable way.

What might be a giant step towards the realisation of this aim, would be to investigate whether these enzymes might be constructed *de novo*, as an essential step in developing a truly combinatorial biosynthesis of complex polyketides.

The 'assembly line' nature of type I polyketide synthases (PKS) that contain sets (called modules) of structurally similar but functionally different enzymatic activities (domains) suggests their potential as a source of "off-the-shelf" enzymatic reagents which can be used to synthesise new and complex polyketide molecules. Outlined below are methodologies for the rapid assembly of DNA units encoding such enzyme domains or modules of enzyme domains.

There are over 40 gene sequences for polyketides that are available from various databases. In addition there are numerous domains known from other synthetic enzymes such as, for example, fatty acid synthase (Joshi and Smith, 1993), peptide synthetases (Elsner *et al.*, 1997) and hybrid polyketide/peptide synthesising enzymes (Paitan *et al.*, 1999; Shen *et al.*, 1999). This amounts to a vast library of domains and modules that cater for a chemical reaction (e.g. stereospecific condensation, dehydration, etc), or in the case of a module, a set of chemical reactions. In order to obtain analogues of a bio-active molecule, research efforts till now have been focused on strategies that involve either chromosomally altering the PKS genes that make the particular molecule (McDaniel *et al.*, 1999) or feeding synthetic intermediates to the PKS (Jacobsen *et al.*,1997) Because of the simplified nature of such experiments, these strategies will remain a fast route towards obtaining a wide variety of drug analogues. However, in the case of compounds like the highly potent anti-cancer discodermolide (TerHaar *et al.*, 1996) the only possible means of obtaining sufficient quantities of the drug is through chemical synthesis. This is because in such cases, the genes responsible for making these bio-active

molecules have not been isolated. The chemical synthesis of large molecules having numerous chiral centres like for example discodermolide, howsoever elegant, is tedious and expensive to scale-up (Marshall and Johns, 1998).

Abbreviations

In addition to those listed in *Biochem. J.* (1986) **233**, 1-24, the following abbreviations have been used:

6-dEB	6-deoxyerythronolide B
6-MSA	6-methylsalicylic acid
6-MSAS	6-methylsalicylic acid synthase
ACP	acyl carrier protein
AT	β-keto acyl transferase
bp	base pair(s) of DNA
DEBS	6-deoxyerythronolide B synthase
DH	β-hydroxyacyl-ACP dehydratase (dehydratase)
ER	enoyl reductase
FAS	fatty acid synthase
kbp	kilobase pair(s)
KR	β-ketoacyl reductase
KS	β-ketoacyl synthase
ORF	open reading frame
PKS	polyketide synthase
RAPS	rapamycin synthase
TE	thioesterase

The Invention

In one aspect the invention provides a method of assembling several DNA units in sequence in a DNA construct. This method comprises the steps of:

a) providing each DNA unit with a restriction enzyme recognition sequence at it's 5' end and with a recognition sequence for the same restriction enzyme at its 3' end that is combined with a recognition site for a DNA modification enzyme,

b) providing a starting DNA construct having an accessible restriction site for the same or a compatible restriction enzyme and cleaving the starting DNA construct with such a restriction enzyme,

c) inserting the desired DNA unit and bringing the ligated product into contact with a DNA modification enzyme such that the restriction site at the 3' end of the inserted DNA unit is abolished,

d) cleaving the ligated product at an accessible unmodified recognition site for the same or a compatible restriction enzyme,

e) repeating steps c) and d) to introduce each desired DNA unit to give a DNA construct containing all the desired units in sequence.

DNA units can be any desired DNA sequence, though usually they encode enzyme domains or modules of two or more enzyme domains. The recognition sequences are usually positioned at the ends of the DNA unit once the DNA unit has been cut with the relevant enzyme, by this it is meant that the recognition sequences are adjacent to the coding sequence, or that they flank the said sequence. An accessible restriction site is herein defined as a restriction site which is unmodified, such that it can be cleaved by a restriction enzyme that normally recognises the sequence of the site. The accessible restriction site is preferably a unique site in the DNA unit or ligated product. Where there is more than one accessible site present, it is possible to perform a partial digest, as known in the art, to obtain digested products in which only the required site is cleaved in the DNA unit. The

DNA modification enzyme employed in the method can be a methylase for example the *dam* methylase of *Escherichia coli*. Other methylases such as *dcm* are also envisaged.

A particular method comprises the steps of

a) providing each DNA unit with an *Xba*I recognition sequence 5'XXTCTAGA3' (where XX is not GA) at it's 5' end and with an *Xba*I recognition sequence 5'GATCTAGA3' at its 3' end.

b) providing a starting DNA construct having an accessible *Xba*I site and cleaving the starting DNA construct with *Xba*I,

c) inserting the desired DNA unit and using a resulting ligated product to transform a dam+ strain of *E. coli*,

d) recovering a resulting plasmid and cleaving the plasmid at an accessible *Xba*I site with *Xba*I,

e) repeating steps c) and d) to introduce each desired DNA unit to give a DNA construct containing all the desired units in sequence.

The recognition sequences for the restriction enzyme and the DNA modification enzyme employed in the method can be created in the DNA units prior to cutting with the restriction enzyme, for example by means of a primer extension reaction. The preferred DNA construct made by the method can be an expression vector capable of facilitating expression of the protein encoded by the desired DNA units.

It is also envisaged that the DNA modification can be removed and the restriction site re-established by replicating the ligated product in a dam- strain of *E. coli* by means of suitable vectors as known in the art.

The invention also encompasses DNA unit assemblies where any given restriction enzyme recognition site can be modified by addition of a certain combination of nucleotide bases in order for it to be protected.

In a further aspect, the invention provides a method of making an assembly of several DNA units in sequence which method comprises the steps of:

a) providing a first DNA unit with a recognition sequence for a first restriction enzyme at its 3' end, and cleaving the said first DNA unit with said first restriction enzyme,

b) providing each other DNA unit with a recognition sequence at its 5' end for a second restriction enzyme which has a compatible ligation sequence with that of the first restriction enzyme, and an upstream recognition sequence for said first restriction enzyme and a downstream recognition sequence for a third restriction enzyme at its 3' end, and cleaving each said other DNA unit with the second and third restriction enzymes,

c) ligating the said first DNA unit with a desired other DNA unit to form a ligated product such that the ligation of the two units abolishes the recognition site for the first restriction enzyme at the ligation junction, and cleaving the ligated product with said first restriction enzyme,

d) ligating the product from c) with a desired DNA unit from b) to form a ligated product and cleaving the ligated product with said first restriction enzyme

e) repeating step d) with each other DNA unit in turn so as to assemble the DNA units in sequence.

A particular method comprises the steps of:

a) providing a first DNA unit with an *Xba*I recognition sequence 5'TCTAGA3' at its 3' end, and cleaving the said first DNA unit with *Xba*I,

b) providing each other DNA unit with a *Spe*I recognition sequence 5'ACTAGT3' at its 5' end, and a downstream *Xba*I recognition sequence 5'TCTAGA3' followed by a downstream *Sma*I recognition sequence 5'CCCGGG3' at its 3' end, cleaving each said other DNA unit with *Spe*I and *Sma*I, and dephosphorylating the 5' end of the cleaved DNA unit,

c) ligating the said first DNA unit with a desired other DNA unit to form a ligated product and cleaving the ligated product with *Xba*I,

d) ligating the product from c) with a desired DNA unit from b) to form a ligated product and cleaving the ligated product with *Xba*I

e) repeating step d) with each other DNA unit in turn so as to assemble the DNA units in sequence.

In one embodiment the assembly can occur via stepwise addition of fragments to a vector.

In an alternative embodiment the first DNA unit can be attached to the solid phase for use in step c). This permits the solid phase to be split and mixed between steps c), d), and e) to make several different assemblies. Methods of attaching DNA units to the solid phase are well know in the art. Preferred solid phase elements are beads attached to the DNA units via a biotinylated nucleotide, as known in the art.

The recognition sequences in one or more of the DNA units are preferably introduced by means of extension primers, as known in the art, though other methods such as the ligation of the required sequences or *in vitro* mutagenesis can also be employed.

The assembly of several DNA units can be inserted into an expression vector and thus used to transform a host capable of expressing the protein encoded by the insert of the vector.

The method is particularly useful where one or more of the DNA units encodes a catalytic or transport protein domain for example a ketoreductase domain from a PKS enzyme or an ACP domain from a hybrid polyketide/peptide synthesising enzyme. Such domains can be derived from enzyme domain DNA sequences from, for example, polyketide synthesising enzymes, peptide synthesising enzymes, hybrid peptide polyketide synthesising enzymes, fatty acid synthesising enzymes or other enzyme domains known in the art.

The DNA units used in the methods of the invention can encode modules comprising one or more catalytic or transport domains. Usually a module contains all of the domains required to complete one condensation step in the synthesis of a target molecule.

Alternative aspects of the invention resulting from the methods of the invention include: DNA constructs or vectors incorporating a DNA assembly

encoding synthetic enzymes, synthetic enzymes encoded by such DNA assemblies, hosts expressing synthetic enzymes, hybrids of transformed hosts expressing synthetic enzymes, and compounds produced by the synthetic enzymes.

Where the product produced by the synthetic enzyme exhibits toxicity to a host stain, this can be worked around e.g. by means of choosing a different strain or mutating the original strain to provide mutants which are more tolerant. The diversity of compounds produced by hosts transformed with the synthetic enzymes of the invention can be further increased by using known methods of using different feedstocks in the fermentation to provide different starter units for the desired product. Where yield of desired synthetic enzyme product is low, routine steps e.g. mutation and selection, can be taken to improve this,

The synthetic enzymes of the invention can also be used in cell-free systems to produce the desired target molecule *in vitro* as known in the art, for example, see Carreras and Khosla (1998).

In a further aspect, the invention provides a method of synthesising a target molecule comprising the steps of

a) examining the composition and stereochemistry of a target molecule,

b) determining which catalytic and transport domains need to be present in a synthetic enzyme in order to catalyse the synthesis of the target molecule,

c) using any one of the methods of the invention to assemble the required DNA units encoding the catalytic and transport domains into a DNA assembly that encodes said synthetic enzyme which is capable of synthesising the target molecule.

d) placing the DNA assembly into a vector to allow expression of the synthetic enzyme in a host capable of synthesising the target molecule after transformation with said vector.

Target molecules are generally bio-active molecules, usually having a predominantly carbon based backbone and usually are macromolecules

comprised of condensed units. The transformed host can be tested for the presence of the target molecule after step d). If yields of the desired compound are low then conventional methods of improving product yield from, for example Streptomycetes, can be employed. Transformed hosts which result from the methods of the invention and their use in producing target molecules are also aspects of the invention. Hosts suitable for transformation with the DNA assemblies of the invention are known in the art and include insect or mammalian cells, though more usually suitable are bacterial cells, for example, the improved host strains described by Ziermann and Betlach (1999).

As stated previously, it is also envisaged that the synthetic enzyme can be used in a cell-free system to produce the target molecule *in vitro*.

A further aspect of the invention is a method of making a synthetic enzyme to catalyse the synthesis of a target molecule comprising the steps of

a) examining the composition and stereochemistry of a target molecule,

b) determining which catalytic and transport domains need to be present in the synthetic enzyme in order to catalyse synthesis of the target molecule,

c) using any one of the methods of the invention to assemble the required DNA units encoding the catalytic and transport domains into a DNA assembly that encodes an enzyme which is capable of synthesising the target molecule.

d) expressing the DNA assembly in a suitable host to produce the enzyme.

In a further aspect the invention provides a library of DNA units encoding catalytic or transport protein domains, wherein each DNA unit has a recognition sequence for a restriction enzyme at it's 5'-end and a second recognition sequence for the same or a compatible enzyme at it's 3'-end which incorporates a recognition sequence for a DNA modifying enzyme.

In a particular embodiment of such a library, each DNA unit has an *XbaI* recognition sequence 5'XXTCTAGA3' (where XX is not GA) at it's 5'-end and an *XbaI* recognition sequence 5'GATCTAGA3' at it's 3'-end

Also provided by the invention is a library of DNA units encoding catalytic or transport protein domains, wherein each DNA unit has a recognition sequence at its 5' end for a first restriction enzyme, and a downstream recognition sequence for a second restriction enzyme followed by a downstream recognition sequence for a third restriction enzyme at its 3' end, such that the DNA units, once restricted by the first and second restriction enzymes can be ligated together to abolish the restriction sites at the ligation junction. In one embodiment of this aspect of the invention each DNA unit has a *SpeI* recognition sequence 5'ACTAGT3' at its 5'-end, and a downstream *XbaI* recognition sequence 5'TCTAGA3' followed by a downstream *SmaI* recognition sequence 5'CCCGGG3' at it's 3'-end

Catalytic or transport protein domains can be derived from any enzyme, for example those listed above. Particularly envisaged are libraries in which the DNA units encode polyketide synthetic domains, comprising two KS domains, at least two AT domains, two KR domains, two DH domains, two ER domains, an ACP domain and a TE domain.

Also provided by the invention are modules comprising a DNA sequence encoding a functional set of polyketide synthetic domains wherein the module has a recognition sequence for a restriction enzyme at it's 5'-end and a second recognition sequence for the same or a compatible enzyme at it's 3'-end which incorporates a recognition sequence for a DNA modifying enzyme. An envisaged module has an *XbaI* recognition sequence 5'XXTCTAGA3' (where XX is not GA) at it's 5'-end and an *XbaI* recognition sequence 5'GATCTAGA3' at it's 3'-end

Alternatively a module comprising a DNA sequence encoding a functional set of polyketide synthetic domains can have a recognition sequence at its 5' end for a first restriction enzyme, and a downstream recognition sequence for a second restriction enzyme followed by a

downstream recognition sequence for a third restriction enzyme at its 3' end, such that the DNA units, once restricted by the first and second restriction enzymes can be ligated together to abolish the restriction sites at the ligation junction. In one particular example, the module has a *Spe*I recognition sequence 5'ACTAGT3' at its 5'-end, and an upstream *Xba*I recognition sequence 5'TCTAGA3' and a downstream *Sma*I recognition sequence 5'CCCGGG3' at it's 3'-end.

Particularly envisaged are modules wherein the DNA units encode polyketide synthetic domains, comprising two KS domains, at least two AT domains, two KR domains, two DH domains, two ER domains, an ACP domain and a TE domain. It is also envisaged that other non-polyketide enzyme domains can be included in the modules provided by the invention.

Also provided by the invention are vectors containing one or more modules. Particularly useful are vectors in which a non-functional recA gene is also present. Such vectors prevent unwanted homologous recombination occurring between domains within the vector upon integration into a suitable host by abolishing the recA gene activity in that host. Thus the invention also provides a method of transforming a host with one or more synthetic DNA assemblies encoding enzyme domains which comprises the steps of:

a) Inserting said DNA assembly into a vector containing a mutated internal fragment of a recA gene sequence such that the vector is capable of undergoing homologous recombination with the recA gene of the host,

b) bringing said vector into contact with a host chromosome under conditions which permit homologous recombination to take place,

c) disrupting the host recA gene by the integration of the DNA of said vector into the chromosome. The expression vector can be used to transform a Steptomyces host. The DNA assemblies contained in the vector can be modules as described herein.

Also envisaged are transformed hosts which prior to transformation with a vector containing one or more modules according to the invention, were already lacking a recA function.

In a further aspect the invention provides kits containing DNA units, DNA modules, vectors, DNA manipulation hosts, DNA modification hosts, expression hosts, or solid phase elements for use in the methods of the invention. For example, one such kit might contain a first DNA unit which is a vector suitable for transforming a suitable host, a library of modules for insertion into that vector, both the first DNA unit and the library having the necessary recognition sites for use in the methods of the invention, together with host strains suitable for the manipulation and expression of the DNA assemblies of the invention.

A *de novo* "domain-by-domain" reconstruction of a hybrid multienzyme from the erythromycin-producing PKS has been achieved by the inventors by assembling DNA units corresponding to the constituent domains. The assembled gene was expressed in *S. erythraea* and the expected compounds were isolated from the bacterial broth. Application of this methodology, or variations of this methodology for making combinatorial assemblies of complex and aromatic PKSs allows for the rapid generation of novel or altered PKS or other synthetic multienzymes and paves the way for a quick and inexpensive synthesis of potentially bio-active molecules.

One alternative to chemical syntheses is to carry out a 'retrobiosynthetic analysis' of the desired molecule, by pinpointing the exact number and type of synthetic enzyme domains that are required for every chemical step, and then assembling the DNA units that encode these enzymes in order to make a hybrid synthetic enzyme. The aim is therefore, to assemble these domains or even modules in a manner as desired, so that the linked enzymes can carry out a progressive synthesis of a desired target molecule. Until now, it has not been possible to find a methodology to assemble these PKS DNA units using restriction enzymes and DNA

ligase to cut and join the DNA pieces together - one of the limiting factors being the non-availability of appropriate restriction enzyme sites in the DNA sequence of the enzymes which synthesise these polyketide drugs. There exist very few unique restriction enzyme sites and even fewer restriction enzymes that do not cut in the polyketide DNA sequence (i.e. are "non-cutters"). However, the restriction enzyme *Xba*I, because of its TA-rich recognition sequence (5'TCTAGA3'), does not cleave the majority of GC-rich polyketide gene clusters. Thus, flanking both ends of the DNA of the desired DNA unit (domain or module) with a recognition sequence that is cleaved on one end by *Xba*I, and on the other end by a restriction enzyme that is compatible with *Xba*I (e.g. *Spe*I) is possible. A vectorial assembly, where such units are progressively joined, leaves one end of the unit that has been constructed by the ligation of *Xba*I and *Spe*I-cut DNA ends, not recognisable by either of the two enzymes, thus making further addition of units possible at only one of the two ends.

This strategy makes use of selective recognition of the restriction enzyme site by the restriction enzyme *Xba*I, depending upon the sequence adjacent to the restriction enzyme site and upon the strain used (dam^+ or dam^-) during the assembly process. The method has been shown to be successful, and by using this methodology to assemble modules, the complete erythromycin-producing PKS (comprising of six modules coded by three large open reading frames) can be built in under 10 days. Even though this time-period is small compared to what it would take to assemble the *ery* PKS genes using conventional methodologies, using a variation of the above mentioned methodology, complete gene-clusters, like the 33 kbp erythromycin PKS, can be built within a matter of hours.

Also described herein, is an approach wherein the assembly of the units itself can also be carried out *in vitro* without the need for an *in vivo* DNA modification step. Furthermore, employing the *in vitro* assembly methodology described below, one is now able to not only construct predetermined PKS genes, but also a randomly constructed combinatorial

library of shuffled domains from one or more known synthetic enzymes. This has immediate and important implications for drug-discovery.

The methodology thus outlined requires DNA units to be modified so that they contain the appropriate 5'and 3' ends (X and X^d respectively). These units are then progressively assembled to achieve the desired gene length. The vector containing the assembled or reconstructed gene is then used to transform an expression system to achieve protein expression. This methodology has been shown to work effectively - the hybrid multienzyme DEBS1-TE was reconstructed by assembling *de novo* the ten constituent domains. The assembled gene, when expressed in *S. erythraea* gave the expected six-membered triketide lactones.

However, in the case of larger molecules like discodermolide, one would require a vectorial assembly of some 50 or so PKS units (if domains). A hypothetical PKS that would make a molecule as large as discodermolide would require 12 modules, each possessing the appropriate KS, AT, ACP and a set of reductive domains (e.g. KR, DH or ER). One would find that some of the domains in this group of 50 would be required to carry out the same catalytic function. For example, if all the hydroxy groups resulting from the ketoreductase activity from all 12 modules are of the same configuration, in effect 12 KRs that function in an identical fashion are required. Also, all 12 ACPs would, of course have the same catalytic function. It would therefore logically be more convenient, and less time-consuming if, to achieve ketoreduction from every one of the 12 modules, one used only one KR domain instead of 12 different ones in all the modules, or one ACP instead of 12 different ACPs. In fact, one can calculate that for *every* possible chemical reaction that can be carried out using PKS domains, one requires a set of only 12 domains, that in theory can be used repeatedly (Figure 1).

It is possible that inter-modular recombination events within the reconstituted PKS or other synthetic enzyme gene, may preclude the use of identical PKS or other enzyme domain DNA units in a set of modules. It

might be expected that, for example (Figure 2) the ACP* DNA in module 1 to recombine with the identical ACP* DNA in module 3. This event can take place, for example, when the expression vector that possesses the assembled gene containing numerous identical PKS DNA units is used to transform a streptomyces host for polyketide production.

The inventors have developed a strategy that can circumvent this problem, therefore making it possible to construct large synthetic enzyme gene clusters using identical domains or modules *repeatedly*. This translates into a less expensive route towards synthetic enzyme gene construction (one would not require to have a start-up library of 200 or so to cover all possibilities), as the set of 12 domains, or similar functional arrangements of domains, are true "off-the-shelf" components for the assembly of PKS genes or genes for other hybrid synthetic enzymes.

The inventors provide methods of DNA assembly that pave the way for a cheap and fast synthesis of a host of bio-active molecules, e.g. the anti-cancer drug Discodermolide.

The examples that follow are better described with reference to the following figures:

Figure 1 shows the chemical/stereochemical choices that each PKS domain can make. A total of 12 domains are required for every conceivable polyketide reaction.

Figure 2 shows integration of a plasmid containing more than one identical DNA unit (ACP*). After the plasmid has integrated in the streptomyces host through homologous recombination with TE, internal recombination can occur to yield truncated PKS genes. This is because the host is recA$^+$.

Figures 3A and 3B show a schematic representation of the assembly process. The *de novo* construction of DEBS1-TE. DNA fragments (units) encoding for the constituent domains of the multienzyme DEBS1-TE were inserted sequentially into the expression plasmid pCJR24.

The final plasmid pAR10 was then expressed in *S. erythraea/JC2* to yield the expected triketide lactone products that are synthesised by the schematically shown re-assembled DEBS1-TE synthase. The amino acid changes made within the linker regions between domains are shown below the actual amino acid sequence. Construction of the expression plasmid pAR10 and structural characterisation of the two triketide lactones shown in the above figure is described in the *methods* section. X – X*ba*I restriction enzyme recognition sequence (5'**TCTAGA**3'), X^d – X*ba*I and Dam methylase recognition sequence (5'**GATCTAGA**3')

Figure 4 shows the methodology of the assembly of DNA units using X*ba*I/dam methylase technology. During the second last stage of assembly, indicated as transform and cut in the figure, transformation of a Dam$^-$strain with plasmid (as it is a dam$^-$strain, even X^d would be cleaved by X*ba*I) is effected. Cutting is achieved by X*ba*I and the DNA unit purified on a gel.

Figure 5 shows the procedure for the assembly of DNA units using X*baI*/dam methylase technology.

Figure 6 shows how an X*ba*I site can be made sensitive to methylation. The RE cuts at the sites shown by arrows. The boxed sequence is methylated in a dam$^+$strain thereby altering the X*ba*I recognition site. The sequence however is not methylated in a dam$^-$strain, and so can still be cleaved by X*ba*I. The X*ba*I recognition sequence (5'**TCTAGA**3') can therefore be selectively cleaved by X*ba*I. Assembly of DNA units uses only one restriction enzyme – X*ba*I.

Figure 7 shows the methodology of the *in vitro* assembly of DNA units – I using solid phase beads with the enzymes X*ba*I, S*pe*I and S*ma*I (other X*ba*I – compatible REs may be used).

Figures 8 and 9 show how the methodology of the *in vitro* assembly of DNA units – II would proceed to the point of placing the DNA assembly into an expression vector for transforming and appropriate host. *In vitro* assembly of DNA units (domains) from the first multienzyme of erythromycin – producing PKS.

Figure 10 shows how in one single ligation, 16 ongoing assemblies are generated. This cascade can obtain exponential proportions. The gene library can be increased by increasing the diversity of the incoming unit.

Figure 11 shows the integration of an expression plasmid into a streptomyces host, using a mutated internal fragment of the recA gene as the region for homologous recombination. The resulting PKS gene can now contain more than one identical DNA units as the strain has been made recA minus.

Figure 12 shows the assembled PKS recADEBS1-TE. The second module is composed of domains that normally belong to the first module.

Figure 13 shows the amino acid sequence alignment of the recA protein of S. lividans (S.l.) and S. ambofaciens (S.a). Percent similarity: 96.496, percent identity: 95.418. Match display thresholds for the alignment(s):

| = identity

: = 2

. = 1

Figures 14A and 14B show a DNA sequence alignment of the recA gene S. lividans (S.l) and S. ambofaciens (S.a). Start of the gene is from 'ATG' and stop is 'TGA'. Percent similarity: 94.713, percent identity: 94.713.

Figure 15 shows how an *XbaI/SpeI* system might be used instead of an *XbaI*/dam methylase system to assemble DNA units, a strategy involving compatible restriction enzymes flanking either end of a DNA unit. An example

of compatible REs would be *Xba*I and *Spe*I. The recognition sequence of *Xba*I is – 5'**TCTAGA**3' and that for *Spe*I is 5'**ACTAGT**3'. After *Xba*I and *Spe*I have cleaved the DNA at their respective sites, the DNA unit can be ligated together as the overhanging is complementary. The junction where any two units are joined is now recognised by either *Xba*I or *Spe*I.

Figure 16 is a schematic representation of the compatibility of *Xba*I- and *Spe*I-digested DNA overhangs. It shows the compatibility of the sticky ends produced by *Xba*I and *Spe*I and how re-ligation abolishes both sites.
Figure 17 shows a schematic representation of the erythromycin-producing polyketide synthase; primary organisation of the genes and their corresponding protein domains. The multienzymes deoxyerythronolide B synthase 1 (DEBS1), DEBS2 and DEBS3 each have two modules, each of which processes one cycle of polyketide chain extension. Each of the six modules is constituted by covalently-linked enzymatic domains. Exploitation of such an enzymatic hierarchy as "of-the-shelf" reagents can lead to synthesis of important chemical compounds.

Figure 18 shows the structure of the anticancer drug discodermolide (top) and the 'retrobiosynthetic approach' towards synthesising a target molecule (a discodermolide). Such an approach would involve opening up the structure (a.), identifying the number and type of polyketide carbon units that would make the discodermolide carbon skeleton (b.), and choosing the PKS DNA units (modules/domains) responsible for the uptake and subsequent processing of the carbon units (c.).

Figure 19 shows the anti-tumour compound octalactin and the strategy behind the retrobiosynthetic approach towards synthesising bio-active molecules. The strategy comprises the steps of:

Identify polyketide units – e.g. whether acetate, propionate, etc,

Break-up and identify – break up the carbon skeleton and identify how many such carbon units are present. Eight units would mean one requires eight modules to make a PKS.

Choose – choose the modules or domains that would be required, form an existing library of such PKS modules and domains.

Assemble – assemble the DNA units (modules/domains/using the invention.

Express – express the assembled gene in a host and check for compound production.

Figure 20 shows a schematic representation of they hypothetical polyketide synthase for synthesising octalactin B, assembled from enzyme units that belong to various PKSs in the public domain.

Figure 21 shows a schematic representation of the hypothetical decarestrictine polyketide synthase for synthesising the anti-cholesterol compound decarestrictine J, assembled from enzyme units that belong to various PKSs in the public domain.

Examples

Example 1: Vectorial assembly of DNA units

DNA units that are to be assembled contain the *Xba*I recognition sequence at either end of the unit. At one of the ends, two nucleotides (GA) are arranged at the 5' end of the *Xba*I recognition sequence (thus making it 5'**GA**TCTAGA3'). This is achieved by first incorporating the *Xba*I recognition sequences in the oligonucleotide primers and then amplifying the desired DNA unit by PCR. The PCR products are then ligated to a pUC-18 vector, used to transform a dam$^+$ strain of *E. coli,* and the clones isolated and sequenced for possible errors in the PCR products. A dam$^+$ strain of *E. coli* - like DH10B™ - methylate the nucleotide **A** in the sequence G**A**TCTAGA, as 5'GATC3' is a sequence that is recognised by the product of the Dam methylase gene (Fujimoto *et al.*,1965; Geier *et al.*, 1979). This makes only one end of the DNA unit cleavable by *Xba*I. The vector is then used to transform a dam$^-$ strain of *E. coli* (e.g. ET12567 - MacNeil *et al.* (1992)) and the plasmid DNA isolated. This DNA is now cleavable at *both* ends of the DNA unit by *Xba*I. When a library of units has been constructed using this strategy, and both ends of these units have been cleaved by *Xba*I, they are progressively inserted into a vector that has a unique *Xba*I site and the ligated products are used always to transform a dam$^+$ strain of *E. coli*, thereby making sure that one end of the DNA unit is always protected from cleavage by *Xba*I through methylation. When the assembly of such units is completed, the final plasmid is integrated into a streptomyces strain for the production of the desired polyketide.

Using this methodology, the polyketide synthase DEBS1-TE, a multienzyme that has the first of the three bimodular erythromycin DEBS enzymes (DEBS1), fused with the erythromycin thioesterase (Cortés *et al.*, 1995) was constructed in a *de novo* fashion. The ten inherent PKS domains in DEBS1-TE, namely, loading module (itself composed of an AT and an ACP), KS1 (ketosynthase of module 1), AT1, KR1, ACP1, KS2

(ketosynthase of module 2), AT2, KR2, ACP2 and TE function in conjunction to catalyse the synthesis of (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-hexanoic acid δ-lactone (**2**), figure 3.

The DNA for all ten domains was amplified by PCR to incorporate the two aforementioned recognition sequences for *Xba*I (5'TCTAGA3' and 5'GATCTAGA3') at the 5' and 3' ends of the DNA unit respectively. The PCR products were cloned in pUC18 vector, sequenced, and then used to transform the dam$^-$ *E. coli* ET12567 strain. To initiate the assembly process, the DNA unit for TE was inserted into *S. erythraea* expression vector pCJR24 (Rowe *et al.*, 1998) which has a unique *Xba*I site. This vector also contains a thiostrepton-resistance gene as a marker for identifying successful integrands. The ligated products were used to transform the dam$^+$ *E. coli* DH10B^{TM} strain and the plasmid DNA isolated. This plasmid (pAR1) can only be singly cleaved with *Xba*I, despite possessing two *Xba*I recognition sequences, as one of the sites (situated at the 3' end of the TE unit) has been methylated by the *E. coli* Dam methylase. The next DNA unit (ACP2 from module 2 of DEBS1) was then ligated to the *Xba*I-cut pAR1, the ligation mixture used to transform DH10B cells and the plasmid DNA isolated. Likewise, the other eight DNA units were successively added to pAR1 to finally yield the expression plasmid pAR10 containing the reconstituted DEBS1-TE gene (Figure 3). The junctions where these domains were joined were chosen in the linker regions that lie between these domains, so as to cause minimum disturbance of the structural features of these domains, that might in turn affect the proficiency of the domains themselves (Figure 3). Plasmid pAR10 was then used to transform *S. erythraea/JC2* - a mutant strain of the wild-type *S. erythraea NRRL2338* that lacks the DEBS genes except for the TE DNA fragment (Rowe *et al.*, 1998). Thiostrepton-resistant colonies were selected upon integration of the vector into the *S. erythraea* chromosome. Single transformants were grown on selective media, as described in the *methods* section. The fermentation broth was extracted with ethyl acetate

and a sample of the organic extract was analysed by gas chromatography-mass spectroscopy (GC-MS). Two peaks were observed, corresponding to molecular massess 158 and 172, indicating the presence of the expected acetate- and propionate- derived polyketides (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-pentanoic acid d-lactone (**1**) and (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-hexanoic acid d-lactone (**2**). Both compounds were isolated and fully characterised by high-pressure liquid chromatography (HPLC), ^{1}H 1D and 2D NMR, ^{13}C NMR, FT-ICR spectrometry, and by comparison with a synthetic standard of (**2**) (Brown *et al.*, 1995). One litre of fermentation broth produces 24 mg of (**1**) and 56 mg of (**2**) - yields that are comparable to those reported elsewhere (Lau *et al.*, 1999). It can therefore be asserted that the ten newly constructed inter-domain junctions have not in any way dimmed the catalytic proficiency of the DEBS1-TE synthase.

In the absence of any crystal-structure data on PKS domains, all genetic engineering efforts known in the art have been based on trial-and-error methods of experimenting with where to join two such domains. As a result, the yield of the synthesised polyketide products have varied depending upon the position in the polypeptide chain at which the domains or modules have been linked (McDaniel *et al.*, 1999; Ruan *et al.*, 1997). The successful functioning of the reconstructed polyketide synthase described above has supplied new information about the inter-domain junction sites. Using this information, and the described methodology for the rapid assembly of these enzyme units, it is now possible to carry out a 'retrobiosynthetic analysis' of target molecules and then to use polyketide and other biosynthetic enzyme domains as truly 'off-the-shelf' reagents to achieve a stereospecific synthesis. There is also the possibility of using this methodology for randomly combining DNA units that encode catalytic e.g. DH or transport e.g. ACP protein domains to generate combinatorial libraries of hybrid synthases. By using a suitable assay system to test for biological activity of the compounds that are generated by such means, it is

possible to go back and isolate the hybrid synthetic gene resposible for the production of these compounds.

From 6-methylsalicilic acid to maitotoxin, nature displays a staggering diversity in compounds that are synthesised by means of 'combinatorial gene-shuffling'. This methodology, or variations of this methodology can be used as effective tools towards harnessing the combinatorial potential of discrete enzymatic units or their sets that are the feature of multi-functional PKS and other systems.

A similar system to the *XbaI/dam* system described above, uses the restriction enzyme *FokI* which has the recognition site:

$$5'GGATG(N)_9\downarrow 3'$$
$$3'CCTAC(N)_{13}\uparrow 5'$$

with the *dcm* methylase of *E.coli.* Adding CCA or CCT to the 5' end of the *FokI* recognition site would make the site dcm sensitive. Furthermore, if the sequence TCTAGA were inserted into the redundant section of the *FokI* restriction site, then the enzyme could be used to generate '*XbaI*-cut ends'.

Methods

E. coli dam$^+$ DH10B^{TM} strain was purchased from Gibco BRL, USA.. *Pfu* DNA polymerase was purchased from Boeringer, Germany.

Construction of the final expression plasmid pAR10 was carried out in several steps, as follows. The ten PKS DNA units were amplified by PCR using *pfu* DNA polymerase. The respective regions of *eryAI* gene, as well as the oligonucleotides used for each PCR are outlined:

LM - segment of *eryAI* gene (Bevitt *et al.*, 1992) extending from nucleotide (N) 588 to N 2389;

5'GG**CATATG**GCGGACCTGTCAAAGCTCTCCGACAGT3' and

5'GG**TCTAGATC**CCAGCCGCGGTCGGTCGGCAGTCCCG3',

KS1 - segment of *eryAI* gene extending from N 2384 to N 3769;

5'GG**TCTAGA**CTCGCTGTTCCACCCCGACCCCACGCGCTCGGGCACC GCGCACCA3' and

5'GGT**CTAGATC**GCGCAGCGCGGCGGACTCGTCGACGGGGGCGAAGGCGG3',

<u>AT1</u> - segment of *eryAI* gene extending from N 3764 to N 4813;

5'GGT**CTAGA**CGGTCTCGCGACGGGAAACGCCGACGGTGCCGCCGTTGGAA3'

and

5'GGT**CTAGATC**CACCGCGACACCGGCGGCGAACGCGCGGGAGAGCGCTTCGC3',

<u>KR1</u> - segment of *eryAI* gene extending from N 4808 to N 6316;

5'GGT**CTAGA**GTCGGTGCACCTGGGCACCGGAGCACGCCGGGTGCCCTT3'

and

5'GGT**CTAGATC**GTCGAAGAGCCTGGTCGGGCGCTGCGCGGTGTA3',

<u>ACP1</u> - segment of *eryAI* gene extending from N 6311 to N 6679;

5'GGT**CTAGA**CGACGCGCGGCGGGCTGCGCCGCAGGCGCCGGCCGAACCGCGGG3'

and

5'GGT**CTAGATC**GGCCGTGG-TCGCCGGTGCCGCCTGCTCGGCT3',

<u>KS2</u> - segment of *eryAI* gene extending from N 6674 to N 8200;

5'GGT**CTAGA**CGAGCCGATCGCGATCGTCGGCATGGCGTGC-CGGCTGC3'

and

5'GGT**CTAGATC**GTGCACGGCCTCGGCGGTGTCGGCGGCGAGC-ACCGCGGCCCGCTCCTC3',

<u>AT2</u> - segment of *eryAI* gene extending from N 8195 to N 9340;

5'GGT**CTAGA**GGCGGTGGCCGACGGCGCGGTGGTT3'

and

5'GGT**CTAGATC**GTCACGAGGGGTGGTGCGGTCCGGCAGCAGCCAGAA3',

<u>KR2</u> - segment of *eryAI* gene extending from N 9335 to N 10639;

5'GGT**CTAGA**CGGCTGGTTCTACC-GGGTCGACTGGACCGAG3'

and
5'GG**TCTAGATC**CGGCCGGGGCCGGGCGGCGG-TGTAGGACT3',
<u>ACP2</u> - segment of *eryAI* gene extending from N 10634 to N 10966; 5'GG**TCTAGA**CCGCATCGTCACGACCGCGCCGAGCGA3'
and
5'GG**TCTAGATC**G-GCGTCGAGGAAA3',
<u>TE</u> - segment of *eryAIII* gene (Donadio *et al.* 1991) extending from N 8753 to N 9602; 5'GG**TCTAGA**CAGCGGGACTCCCGCCCGGGAAGCG3'
and
5'GGGCTAGC**TCTAGATC**ATGAATTCCCTCCGCCCAGCCAGGCGTC3'.

All PCR products were 5' phosphorylated and ligated to *Sma*I-cut, dephosphorylated pUC18 vector and used to transform *E. coli* DH10B electrocompetent cells. The desired plasmids - containing the amplified DNA fragments were isolated and sequenced using standard pUC forward and reverse primers. No mistakes in the amplified products were detected. All ten plasmids were then used to transform the *E.coli* ET12567 dam$^-$ strain. Isolated DNA was digested with *Xba*I restriction enzyme and desired fragments isolated and purified. The TE unit was then ligated to *Xba*I-cut pCJR24 vector and the ligation products used to transform *E. coli* DH10B electrocompetent cells. Plasmid pAR1 was isolated, digested with *Xba*I, and ligated to the ACP2 fragment, and ligation products treated as mentioned above. The other DNA fragments, namely, KR2, AT2, KS2, ACP1, KR1, AT1 and KS1 were sequentially added to finally yield plasmid pAR10. This plasmid was then digested with *Nde*I and *Xba*I restriction enzymes and ligated with the LM fragment previously digested with the same two enzymes. The ligated products were used to transform *E. coli* DH10B electrocompetent cells and the final expression plasmid pAR10 isolated. Plasmid pAR10 was then used to transform *S. erythraea/JC2* strain and colonies carrying the expression plasmid were selected through resistance to thiostrepton upon integration of the plasmid into the S. *erythraea* chromosome. Single transformants were picked and grown on

tap-water medium plates supplemented with thiostrepton, following which single transformants were grown in 5X200ml of SM3 liquid media supplemented with 5 ug/ml of thiostrepton for seven days (Rowe *et al.*, 1998). Cells were removed by centrifugation, the supernatant was saturated with NaCl and extracted three times with equal volumes of ethyl acetate at pH 4.0. The solvent was evaporated to yield 1.12 g of crude product. A sample of this crude product was analysed by GC-MS. Two peaks were observed, corresponding to molecular masses 158 and 172, indicating the presence of the expected acetate- and propionate- derived polyketides (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-pentanoic acid δ-lactone (**1**) and (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-hexanoic acid δ-lactone (**2**). Compounds (1) and (2) were found to be structurally identical to those reported previously (Cortés *et al.*,1995).

*Characterisation of (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-pentanoic acid δ-lactone (**1**)*

^{1}H NMR ($CDCl_3$, 500 MHz) $^{\delta}$H 4.45-4.35 (1H, dq, J = 6.56 and 1.62 Hz, C_5-H), 3.8 (1H, dd, J = 10.15 and 4.17 Hz C_3-H), 2.45-2.70 (1H, br, O-H), 2.42 (1H, dq, J = 10.0 and 6.97 Hz C_2-H), 2.05 (1H, m, C_4-H), 1.37 (3H, d, J = 7.17 Hz, C_2-CH_3), 1.32 (3H, d, J = 6.74 Hz, C_5-CH_3), 0.95 (3H, d, J = 7.20 Hz, C_4-CH_3) ppm. ^{13}C NMR ($CDCl_3$, 250 MHz) δ 174.20, 76.15, 73.62, 39.42, 38.14, 18.11, 14.24, 4.48.

*Characterisation of (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-hexanoic acid δ-lactone (**2**)*

^{1}H NMR ($CDCl_3$, 500 MHz) dH 4.13 (1H, ddd, J = 8.12, 5.93 and 2.19 Hz, C_5-H), 3.82 (1H, m, C_3-H), 2.42-2.50 (1H, dq, J = 10.17 and 7.08 Hz, C_2-H), 2.12-2.19 (1H, m, C_4-H), 1.77-1.86 (1H, m, one of C_6-H_2), 1.52-1.61 (1H, m, one of C_6-H_2), 1.4 (3H, d, J = 7.09 Hz, C_2-CH_3), 1.0 (3H, t, J = 7.42 Hz, C_6-CH_3), 0.97 (3H, d, J = 6.96 Hz, C_4-CH_3) ppm. ^{13}C NMR ($CDCl_3$, 250 MHz) d 173.56, 81.34, 73.96, 40.08, 36.76, 25.27, 14.27, 9.88, 4.37.

<u>Example 2: *in vitro* assembly of DNA units</u>

Figure 7 outlines the strategy for the *in vitro* assembly of PKS DNA units. The inventors have constructed the multienzyme DEBS1-TE. The *in vivo* construction of the gene for DEBS1-TE, it should be noted, took 12 days to complete. The *in vitro* assembly on the other hand was completed in 2 days.

All ten domains, namely, LM, KS1, KR1, AT1, ACP1, KS2, AT2, KR2, ACP2 and TE were amplified by means of PCR. The forward primer in all cases, except the LM contained the *Spe*I recognition sequence 5'ACTAGT3' while the reverse primer was engineered in such a way that it contained the *Xba*I recognition sequence 5' TCTAGA3' and *Sma*I recognition sequence 5'CCCGGG3' downstream of the *Xba*I site (Figure 7). The amplification of the LM was carried out using a biotinylated forward primer and a reverse primer that contained the *Xba*I recognition sequence (5'TCTAGA3'). All the PCR products were cloned in pUC-18 vector and the resulting plasmids sequenced to detect possible errors introduced by PCR. All plasmids, except the one containing the LM unit were then digested with *Spe*I and *Sma*I, dephosphorylated in order to remove the 5' phosphate group and the appropriate fragments isolated and eluted. The LM unit was cleaved with *Xba*I and attached to a bead that was coated with streptavidin (following the manufacturer's instructions) as shown in figure 7.

The assembly process was initiated by adding DNA ligase to the tube containing a large excess of the first unit (KS1) and LM-bead. The reason for having a large excess of the KS1 unit compared to the LM-bead unit is to favour the LM-bead ligating to the incoming unit, as opposed to the self-ligation of the LM-bead (see figure 7). The ligation of the two DNA fragments is unidirectional as only the *Spe*I-cut end of KS1 complements the *Xba*I-cut end of the LM-bead. After the ligation was complete, the desired product of the ligation reaction, namely 'bead-LM-KS1' was separated from the reaction mixture and washed. This product was then cleaved with *Xba*I, in order to activate the 3' end of KS1. The beads were washed again to remove the small *Xba*I-*Sma*I DNA fragment that was

released from the 3' end of KS1 as a result of RE cleavage. The 'activated' bead-LM-KS1 unit was then ligated with *Spe*I, *Sma*I-cut and 5' dephosphorylated AT1. The *Spe*I-cut 5' end of AT1 complemented the *Xba*I-cut 3' end of KS1 to give bead-LM-KS1-AT1 as shown in figure 8. This product was separated from the reaction mixture and washed as before. The 3' end of AT1 in this product was then 'activated' through cleavage by *Xba*I, and the assembly process continued.

Finally, *Spe*I, *Sma*I-cut and 5' dephosphorylated TE unit was ligated with the DNA fragment that was now bead-LM-KS1-AT1-KR1-ACP1-KS2-AT2-KR2-ACP2 as shown in figure 9. The 3' end of the latter fragment was 'activated' by digesting it with *Xba*I. The assembled DEBS1-TE gene was then inserted in the expression plasmid pCJR24 and the resulting plasmid used to transform a streptomyces strain. The expected triketide lactone products were isolated and structurally characterised.

Use of the *in vitro* technology described above drastically reduces the time it takes to assemble predetermined or randomly shuffled genes. Also, the possibility of continuing with the assembly process while having numerous different assembly arrays attached to the beads, and splitting and mixing the beads between each unit/module addition from a library of units/modules, results finally in the generation of a cascade of different assemblies (Figure 10). These assembled genes can then be cloned simultaneously and expressed in a suitable host. An assay system can then be used to identify those assembled genes that yield bio-active compounds.

Example 3: Retrobiosynthetic synthesis of a target molecule

A strategy employing the invention in order to construct the highly potent anti-breast cancer drug discodermolide, the anticholesterol compound decarestrictine, and the antitumour compound octalacin using polyketide synthase domains/modules is outlined below.

Discodermolide

The drug discodermolide (Figure 18), isolated from the marine sponge '*Discodermia disoluta*', has been identified as a highly potent anti-cancer compound and 80 times more effective than the well known anticancer drug Taxol (TerHarr *et al.*, 1996). It has the same mechanism of action as Taxol, even though it is structurally different from the latter.

One can infer from its structure (Figure 18) that discodermolide is a polyketide and can therefore be constructed from a system that has the basic enzymatic building blocks (domains and modules) that make other polyketides like erythromycin and rapamycin. Having predicted that approximately 45 domains housed in 12 modules would be required in order to carry out the chemistry that accounts for the functionalities on the carbon skeleton of discodermolide, one can now begin to construct such a system. All one has to do is to identify the type and nature of the domains/modules that one requires to generate the observed functionalities, and then assemble these units in the desired order (Figure 18). The resulting DNA assembly can then be put into a bacterial strain that makes a functional polyketide synthase.

Until now, it would have been exceedingly difficult, if not impossible to assemble 45 or so pieces of DNA in the wanted order, for several reasons. Firstly, one would have to look for two different restriction enzymes every time one needed to assemble two DNA segments. This is because if one uses just one restriction enzyme at either end of the

domain, the already-assembled piece/pieces of DNA would be cleaved from the assembly every time one decided to insert a new domain. Secondly, in GC-rich DNA like the polyketide synthase producing *Streptomyces* strain, unique restriction enzyme sites are few and far between. To a molecular biologist, the task of assembling 40 pieces of DNA with the limitations mentioned above, would seem an insurmountable one. One would rather attempt to isolate the genes that make the drug at the first place than consider carrying out "step-by-step" reconstruction of the gene itself. In the case of discodermolide, even the last possibility is in the realms of fantasy. The organism within the marine sponge that makes the drug has not been identified. The only way discodermolide can be made available is through chemical synthesis - there have been a few chemical routes reported in literature recently (Marshall and Johns, 1998 and references therein). However, as is the case with most other complex molecules, large scale production of discodermolide, using the chemical route would turn out to be outrageously expensive. Chemists have been using the retrosynthetic analysis approach towards total synthesis of important bioactive molecules. This approach breaks the target compound into many smaller pieces - easily synthesised - which are then re-assembled.

The type of polyketide or other synthetic enzyme domains required in order to construct the target molecule from the starting units are identified using a "retrobiosynthetic analysis" approach for discodermolide,

by matching which molecules need to be condensed to form the macromolecule with the enzyme domains that carry out the required catalysis to build the macromolecule.

Having identified the enzyme units that are required, the unit-DNA segments are amplified using the polymerase-chain-reaction (PCR) - from the library of existing polyketide synthase unit-DNA, and the appropriate recognition sequences are attached to each unit-DNA fragment. All of the unit fragments are then replicated in a dam$^-$ strain whereby both the unmodified and modified sequences (5'TCTAGA3' and 5'**GA**TCTAGA3' respectively) are cleaved by the restriction enzyme *Xba*I.

Having constructed this library of appropriate PKS or other synthetic enzyme units, the corresponding DNA units are then assembled. The assembled DNA piece is then placed in a vector, so that it can be inserted in a bacterial strain to yield the desired synthetic protein. Suitable vectors have an antibiotic resistance marker (for selection of this vector on an antibiotic-rich media) and an "origin-of -replication" (ori). Ori is essential for the independent growth of the vector in any strain. Particularly suitable vectors for the expression of the synthetic enzymes of the invention are the actinomycete vectors described by Rowe *et al.* (1998).

The strain is then grown in a media that is supplemented with the antibiotic, the resistance gene for which is present in the vector.

Figures 4 and 5 show how the assembly proceeds. The first domain is inserted into a vector that is cut by cleavage with *Xba*I. After the ligation

of the domain has taken place with the vector, the DNA is put in a bacterial strain that is dam$^+$ and grown. Finally, bacterial colonies that have the desired vector-domain DNA are identified and DNA isolated from them. The whole procedure is cheap and fast. Only one restriction enzyme (*XbaI*) is made use of, routine cloning technology is employed, the desired DNA fragment is obtained, which can then be expressed in a Streptomyces strain to yield the polyketide synthase.

The *in vivo* "domain-by domain" construction of the discodermolide producing polyketide synthase would take approximately 55 days via this method. In comparison, assembly of modules would take less time, as one would need to assemble fewer pieces. Most importantly, once the synthase is shown to be functionally active, a large fermentation of the bacterial strain can be carried out, and the drug isolated in however much quantity one requires - unlike the chemical route where the starting materials have to be freshly synthesised every time one requires the target compound. Employing such a strategy would lead to a quick and inexpensive synthesis of important bioactive molecules like discodermolide.

Retrobiosynthetic analysis

The whole approach (retrobiosynthetic analysis followed by identification of PKS units, followed by assembly of PKS units) is made clearer in the following two examples.

Octalactin

A new addition to the rare class of eight-membered lactone natural products is the family of Octalactin. Octalactin A and B (Figure 20) are natural products isolated from the marine gorgonian octocoral '*Pacifigorgia sp.*' (Tapiolas et. al., 1991). Octalactin A shows very strong cytotoxicity toward B-16-F-10 murine melanoma and HCT-116 human colon tumour cell lines and is a promising drug candidate, while octalactine B displayed no such activity (Tapiolas et. al., 1991). Total syntheses of both octalactin A and B have been reported in literature. One such synthesis (Buszek, et. al., 1994) typically involves more than 12 chemical steps in leading to the target molecules. Clearly, large-scale production of octalactins using chemical synthesis is industrially not viable. On the other hand, the genes that code for the enzymes that make octalactins have not be identified or isolated. This means that at present, modified octalactins can only be made using chemical synthesis. A gene is constructed from the available PKS spare parts - that would code for the enzymes that would make octalactin B. Octalactin B can then be converted into the cytotoxic octalactin A by one-step stereospecific epoxidation. Also, once the gene for octalactin B is constructed and shown to make the octalactin PKS, genetic engineering on this gene would yield modified octalactin PKSs that in turn would synthesise octalactin analogues.

Clearly, a polyketide, the carbon skeleton of octalactin B (Figure 19) can be seen to be assembled by acetate and propionate units. The uptake

and assembly of these units in the prescribed sequence, as well as the functionalities that decorate the carbon chain of octalactin can be assigned to various PKS modules (see figure 19). Once a decision has been made regarding the type and nature of PKS modules, they can be strung together to make a gene using the invention. This gene can then be expressed in a suitable host in order to look for octalactin B production. The retrobiosynthetic approach towards octalactin is shown in detail in figure 19. A choice of what modules to select from the PKS module library is followed by amplification of the modular DNA fragments using the oligonucleotides such that the 5' and the 3' ends of every DNA fragment have the restriction enzyme recognition sites stated under the description of the invention. The choice of modules that, when assembled, would make the 'octalactin gene' is displayed as a schematic representation in figure 20.

Decarestrictine J

The molecule decarestrictine J can be synthesised using the retrebiosynthetic approach. Decarestrictine J is a ten-membered lactone that comes from the family of decarestrictines, shown to display strong anti-cholesterol activity (Grabley et. al., 1992). The total synthesis of Decarestrictine J has been reported and involves numerous chemical steps (Yamada et. al., 1995). The target molecule (figure 21) can be conceived to be formed by assembly of five acetate polyketide units. Using the retrobiosynthetic approach, one can identify the PKS domains/modules that

would be required for the carbon skeleton of decarestrictine J. A hypothetical decarestrcitine PKS is shown in figure 21. The loading module, as well as the four internal modules along with the TE domains can be conveniently assembled using the invention. The assembled 'decarestrictine gene' can then be expressed in a suitable host in order to check for the production of decarestrictine J.

In summary, the retrobiosynthetic approach involves the following steps;

a). Identification of the *number* and *nature* of carbon units that make up the target molecule

b). Identification of the modules/domains from libraries of polyketide/peptide synthetase/fatty acid/etc. encoding units that are responsible for the uptake of the said carbon units and the nature and degree of functionalisation of the carbon chain

c). Assembly of the said modules/domains using the methods of the invention

d). Expression of the assembled gene in a suitable expression host.

Example 4: Transforming strains with DNA encoding similar synthetic enzyme domains

A method for transforming expression strains with DNA encoding similar synthetic enzyme domains has been devised. Instead of using the TE PKS DNA fragment as a region of integration from the assembled gene into a streptomyces host (*S. erythraea*/JC2, Rowe *et al.*, 1998), a mutated *recA* gene fragment from streptomyces is used. The assembly process is carried

out in a *recA⁻* *E. coli* strain (e.g. DH10B) as previously described. As this strain is recA⁻, one can assemble any number of identical DNA units. The vector, into which the assembled gene is being constructed, contains a portion of a streptomyces *recA* gene. This *recA* fragment carries a mutation. After the synthetic enzyme gene has been assembled, the vector is used to transform a streptomyces host (e.g. *S. lividans* or *S. erythraea*). The fragment of *recA* gene carrying a mutation recombines with the *recA* gene of the streptomyces host, abolishing the functional recA gene and making the strain recombination minus (Figure 11). This means that an event, such as the one described in figure 2 is now not possible. The strain is then grown to look for the encoded enzyme product. This strategy is tested by assembling a functional PKS gene having more than one type of identical DNA units (Figure 12).

Construction of the PKS multienzyme recDEBS1-TE

RecA protein has been characterised as a multifunctional enzyme that is essential for homologous recombination, DNA repair, SOS response and DNA rearrangements (Miller and Kokjohn, 1990). Most of the routinely used strains of *E. coli* are *recA⁻*. The gene for *recA* has been identified from many streptomyces strains. The first streptomyces *recA* gene to be characterised and isolated was from *S. lividans* (Nußbaumer and Wohlleben, 1994) *RecA* mutants have since been generated in *S. ambofaciens* (Aigle *et al.*, 1997). The streptomyces *recA* protein has approximately 372 amino acid residues (Figure 13). DNA sequence analysis suggests a coding region of 1122 bp, and is found to be highly conserved within streptomyces (Figure 14). In fact the *recA* mutants of *S. ambofaciens* were generated by integrating a mutated portion of the *S. lividans recA* gene into the *S. ambofaciens* host. It was found that a *recA* mutant lacking 30 aa from the C-terminus of the protein inhibited recombination events in *S. ambofaciens* (Aigle *et al.*, 1997).

A *recA* mutant of the streptomyces host that is used for expression of the assembled gene was generated.

The oligonucleotides:

5'- GG*TCTAGA*ATTCGGCAAGGGCGC**C**GGTCATGCGCAT-3' and

5'- GG*TCTAGATCT*GCGGCGTCGGCCGGGGCGGCGGAGGCG-3'

were used as the forward and reverse primers respectively and the 1000 bp internal region of *S. lividans recA* gene (Nußbaumer and Wohlleben, 1994) was amplified using *pfu* polymerase. An additional nucleotide (**C**) was incorporated into the forward primer to generate a frame shift in the amplified *recA* gene fragment. The PCR product was cloned in pUC-18 vector and sequenced to detect for possible errors during PCR. The 1.0 kbp *recA* fragment, flanked at both ends by an *Xba*I site was then inserted in the expression vector pCJR24 that has a unique *Xba*I site. The ligation mixture was used to transform *E. coli* DH10B cells and the desired plasmid DNA isolated. The resulting plasmid (pA*RecA*24) contains a non-methylated *Xba*I site at the 5' end of the *recA* gene fragment. The ten PKS DNA units, namely, TE, two each of ACP1, KR1, AT1 & KS1, and LM were inserted into the plasmid pA*RecA*24 to finally yield the expression plasmid p*RecA*D1TE. This plasmid was used to transform wild-type *S. lividans* protoplasts, and thiostrepton resistant colonies were grown in defined liquid media as described above. The compound (Figure 12) was isolated from the bacterial broth and chemically characterised.

Thus, it has been shown that a gene carrying interspaced DNA units that are identical in structure as well as function does not lead to internal recombination events, as the native *recA* gene of the streptomyces host has been disrupted. Furthermore, it has been shown that it is possible to use identical domains to reach the objective of generating hybrid synthetic enzyme systems. This strategy will greatly reduce the number of domains that otherwise have to be employed for the purposes of *de novo* PKS gene assembly that yields the desired chemical compounds. The inventors have established a set of 12 domains that are capable of functioning robustly and are independent of flexibility and spacial constraints - problems that beset the choice of domains and modules previously.

References

Aigle, B., Holl, A-C., Angulo, J.F., Leblond, P. and Decaris, B. (1997) Characterization of two *Streptomyces ambofaciens recA* mutants: identification of the *recA* protein by immunoblotting. *FEMS Microbiol. lett.*, **149**, 181-187.

Bevitt, D.J., Cortés, J., Haydock, S.F. and Leadlay, P.F. (1992) 6-Deoxyerythronolide B synthase 2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur. J. Biochem.*, **204**, 38-49.

Brown, M.J.B., Cortés, J., Cutter, A.L., Leadlay, P.F. and Staunton, J. (1995) A mutant generated by expression of an engineered DEBS1 protein from the erythromycin-producing polyketide synthase (PKS) in *Streptomyces coelicolor* produces the triketide as a lactone, but the major product is the nor-analogue derived from acetate as starter acid. *J. Chem. Soc., Chem. Commun.*, 1517-1518.

Buszek, KR., Sato, N. and Jeong, Y.M. (1994) Total synthesis of octalactin-A and octalactin-B. *J. Amer. Chem. Soc.* **116**, 5511-5512.

Carreras C. and Khosla C. (1998) Purification and *in vitro* reconstitution of the essential protein components of an aromatic polyketide synthase. *Biochemistry* **37**,2084-2088.

Cortés, J., Wiesmann, K.E.H., Roberts, G.A., Brown, M.J.B., Staunton, J. and Leadlay, P.F. (1995) Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science*, **268**, 1487-1489.

Donadio, S., McAlpine, J.B., Sheldon, P.J., Jackson, M. and Katz, L. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7119-7123.

Donadio, S., Staver, M.J., Mcalpine, J.B., Swanson, S.J. and Katz, L. (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science*, **252**, 675-679

Elsner, A., Engert, H., Saenger, W., Hamoen, L., Venema, G. and Bernhard, F. (1997) Substrate specificity of hybrid molecules from peptide synthetases. *J. Biol. Chem.* **272**, 4814-4819.

Fujimoto, D., Srinivasan, P.R. and Borek, E. (1965) On the nature of the deoxyribonucleic acid methylases. Biological evidence for the multiple nature of the enzymes. *Biochemistry* **4**, 2849-2855.

Geier, G. E. and Modrich, P. (1979) Recognition sequence of the *dam* methylase of *Escherichia coli* K12 and mode of cleavage of *Dpn* I endonuclease. *J. Biol. Chem,* **254**, 1408-1413.

Grabley, S., Granzer, E., Hutter, K., Ludwig, D., Mayer, M., Thiericke, R., Till, G., Wink, J., Phillips, S. and Zeeck, A. (1992) *J. Antibiot.* **45**, 56-65.

Jacobsen, J.R., Hutchinson, C.R., Cane, D.E. and Khosla, C. Precursor-directed biosynthesis of erythromycin analogs by an engineered polyketide synthase. *Science* **277**, 367-369 (1997)

Joshi, A.K. and Smith S. (1993) Construction of a cDNA encoding the multifunctional animal fatty acid synthase and expression in *Spodoptera frugiperda* cells using baculoviral vectors. *Biochem J.*,**296**, 143-149.

Kao, C.M., Luo, G.L., Katz, L., Cane, D.E. and Khosla, C. (1995) *J. Am. Chem. Soc.*, **117**, 9105-9106.

Kao, C.M., Luo, G.L., Katz, L., Cane, D.E. and Khosla, C. (1996) *J. Am. Chem. Soc.*, **118**, 9184-9185.

Kao, C.M., Luo, G.L., Katz, L., Cane, D.E. and Khosla, C., (1994) *J. Am. Chem. Soc.*, **116**, 11612-11613.

Kuhstoss, S., Huber, M., Turner, J.R., Paschal, J.W. and Rao, R.N. (1996) *Gene*, **183**, 231-236.

Lau, J., Fu, H., Cane, D. E. and Khosla, C. (1999) Dissecting the role of Acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry*, **38**, 1643-1651.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H. and MacNeil, T. (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesisutilizing a novel integration vector. *Gene* **111**, 61-68.

Marsden, A.F.A., Wilkinson, B., Cortés, J., Dunster, N.J., Staunton, J. and Leadlay, P.F. (1998) *Science*, **279**, 199-202.

Marshall, J.A. and Johns, B.A. (1998) Total synthesis of (+)-discodermolide. *J. Org. Chem.* **63**, 7885-7892.

McDaniel, R. et al. (1999) and references therein. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. USA*, **96**, 1846-1851.

Miller, R.V. and Kokjohn, T.A. (1990) General microbiology of *recA*: Environmental and evolutionary significance. *Annu. Rev. Microbiol.*, **44**, 365-394.

Nußbaumer, B. and Wohlleben, W. (1994) Identification, isolation and sequencing of the *recA* gene of *Streptomyces lividans* TK24. *FEMS Microbiol. lett.*, **118**, 51-56.

Oliynyk, M., Brown, M.J.B., Cortés, J., Staunton, J. and Leadlay, P.F. (1996) *Chem. Biol.*, **3**, 833-839.

Paitan, Y., Alon, G., Orr, E., Ron, E.Z., and Rosenberg, E. (1999) The first gene in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase. *J. Mol. Biol.* **286**,465-474.

Rowe, C.J., Cortés, J., Gaisser, S., Staunton, J., Leadlay, P.F. (1998) Construction of new vectors for regulated high-level expression in actinomycetes. *Gene*, **216**, 215-223.

Ruan, X., Pereda, A., Stassi, D.L., Zeidner, D., Summers, R.G., Jackson, M., Shivakumar, A., Kakavas, S., Staver, M.J., Donadio, S. and Katz, L. (1997) Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *J. Bacteriol.* **179**, 6416-6425.

Shen, B., Du, L., Sanchez, C., Chen, M. and Edwards, D.J. (1999) Bleomycin biosynthesis in Streptomyces verticillus ATCC15003: A model of hybrid peptide and polyketide biosynthesis. *Bioorganic Chemistry* **27**, 123-129.

Tapiolas, D.M., Roman, M., Fenical, W., Stout, T.J. and Clardy, J. (1991)

Octalactin-A and Octalactin-B - cytotoxic 8-membered-ring lactones from a marine bacterium, *Streptomyces sp. J. Amer. Chem. Soc.* **113**, 4682-4683.

TerHaar, E., Kowalski, R.J., Hamel, E., Lin, C.M., Longley, R.E., Gunasekera, S.P., Rosenkranz, H.S. and Day, B.W. (1996) Discodermolide, a cytotoxic marine agent that stabilizes microtubules more potently than taxol. *Biochemistry* **35**, 243-250.

Yamada, S., Tanaka, A. and Oritani, T. (1995) Total synthesis of Decarestrictine-J. *Biosci. Biotech. & Biochem.* **59**, 1657-1660

Ziermann, R. and Betlach, M.C. (1999) Recombinant Polyketide Synthesis in *Streptomyces*: Engineering of improved host strains. *BioTechniques* **26**, 106-110.

CLAIMS

1. A method of assembling several DNA units in sequence in a DNA construct, which method comprises the steps of

a) providing each DNA unit with a restriction enzyme recognition sequence at it's 5' end and with a recognition sequence for the same restriction enzyme at its 3' end that is combined with a recognition site for a DNA modification enzyme.

b) providing a starting DNA construct having an accessible restriction site for the same or a compatible restriction enzyme and cleaving the starting DNA construct with such a restriction enzyme,

c) inserting the desired DNA unit and bringing the ligated product into contact with a DNA modification enzyme such that the restriction site at the 3' end of the inserted DNA unit is abolished

d) cleaving the ligated product at an accessible unmodified recognition site for the same or a compatible restriction enzyme,

e) repeating steps c) and d) to introduce each desired DNA unit to give a DNA construct containing all the desired units in sequence.

2. The method of claim 1 wherein the DNA modification enzyme is a methylase.

3. The method of claim 2 wherein the methylase is the *dam* methylase of *Escherichia coli.*

4. The method of claim 3 which comprises the steps of

a) providing each DNA unit with an *Xba*I recognition sequence 5'XXTCTAGA3' (where XX is not GA) at it's 5' end and with an *Xba*I recognition sequence 5'GATCTAGA3' at its 3' end.

b) providing a starting DNA construct having an accessible *Xba*I site and cleaving the starting DNA construct with *Xba*I,

c) inserting the desired DNA unit and using a resulting ligated product to transform a dam+ strain of *E. coli*,

d) recovering a resulting plasmid and cleaving the plasmid at an accessible *Xba*I site with *Xba*I,

e) repeating steps c) and d) to introduce each desired DNA unit to give a DNA construct containing all the desired units in sequence.

5. The method of any one of claims 1 to 4, wherein the recognition sequences for the restriction enzyme and the DNA modification enzyme are created in the DNA units prior to cutting with the restriction enzyme.

6. The method of claim 5 wherein the restriction sites are created in the fragment by means of a primer extension reaction.

7. The method of any one of claims 1 to 6, wherein the DNA construct is an expression vector capable of facilitating expression of the protein encoded by the desired DNA units

8. The method of claim 3 or claim 4, wherein the DNA modification is removed and the restriction site re-established by replicating the ligated product in a dam- strain of *E. coli* by means of a suitable vector.

9. A method of making an assembly of several DNA units in sequence which method comprises the steps of:

a) providing a first DNA unit with a recognition sequence for a first restriction enzyme at its 3' end, and cleaving the said first DNA unit with said first restriction enzyme,

b) providing each other DNA unit with a recognition sequence at its 5' end for a second restriction enzyme which has a compatible ligation sequence with that of the first restriction enzyme, and a downstream recognition sequence for said first restriction enzyme followed by a downstream recognition sequence for a third restriction enzyme at its 3' end, and cleaving each said other DNA unit with the second and third restriction enzymes,

c) ligating the said first DNA unit with a desired other DNA unit to form a ligated product such that the ligation of the two units abolishes the recognition site for the first restriction enzyme at the ligation junction, and cleaving the ligated product with said first restriction enzyme,

d) ligating the product from c) with a desired DNA unit from b) to form a ligated product and cleaving the ligated product with said first restriction enzyme

e) repeating step d) with each other DNA unit in turn so as to assemble the DNA units in sequence.

10. The method of claim 9 which method comprises the steps of:

a) providing a first DNA unit with an *Xba*I recognition sequence 5'TCTAGA3' at its 3' end, and cleaving the said first DNA unit with *Xba*I,

b) providing each other DNA unit with a *Spe*I recognition sequence 5'ACTAGT3' at its 5' end, and a downstream *Xba*I recognition sequence 5'TCTAGA3' followed by a downstream *Sma*I recognition sequence 5'CCCGGG3' at its 3' end, cleaving each said other DNA unit with *Spe*I and *Sma*I, and dephosphorylating the 5' end of the cleaved DNA unit,

c) ligating the said first DNA unit with a desired other DNA unit to form a ligated product and cleaving the ligated product with *Xba*I,

d) ligating the product from c) with a desired DNA unit from b) to form a ligated product and cleaving the ligated product with *Xba*I

e) repeating step d) with each other DNA unit in turn so as to assemble the DNA units in sequence.

11. The method of claim 9 or claim 10 wherein the assembly occurs via stepwise addition of fragments to a vector

12. The method of claim 9 or claim 10 wherein the said first DNA unit is attached to the solid phase for use in step c)

13. The method of claim 12, wherein the solid phase is split and mixed between steps c), d), and e) to make several different assemblies.

14. The method of any one of claims 9-13, wherein the recognition sequences in one or more of the DNA units are introduced by means of extension primers.

15. The method of any one of claims 9-14 wherein the assembly of several DNA units is inserted in to an expression vector which is used to transform a host capable of expressing the protein encoded by the vector

16. The method of any one of claims 1-15, wherein one or more of the DNA units encodes a catalytic or transport protein domain. *(see Kleinkauf peptide/polyketide systems paper)*

17. The method of claim 16 wherein one or more of the DNA units are derived from polyketide synthesising enzyme domain DNA sequences.

18. The method of claim 16 wherein one or more of the DNA units are derived from peptide synthesising enzyme domain DNA sequences.

19. The method of claim 16 wherein one or more of the DNA units are derived from hybrid peptide polyketide enzyme domain DNA sequences.

20. The method of claim 16 wherein one or more of the DNA units are derived from fatty acid synthesising enzyme domain DNA sequences

21. The method of claim 16 wherein one or more of the DNA units encode modules comprising one or more catalytic or transport domains

22. DNA constructs incorporating one or more DNA assemblies encoding synthetic enzymes made by any one of the methods of claims 1-21.

23. Synthetic enzymes encoded by one or more DNA assemblies made by the methods of anyone of claims 1-21

24. Hosts expressing DNA constructs encoding one or more synthetic enzymes made by any one of the methods of claims 1-21.

25. Hybrids of transformed hosts expressing one or more DNA constructs encoding synthetic enzymes incorporating a DNA assembly made by any one of the methods of claims 1-21.

26. Compounds produced by synthetic enzymes encoded by DNA assemblies made by any one of the methods of claims 1-21.

27. A method of synthesising a target molecule comprising the steps of

a) examining the composition and stereochemistry of a target molecule,

b) determining which catalytic and transport domains need to be present in a synthetic enzyme in order to catalyse the synthesis of the target molecule,

c) using any one of the methods of claims 1-21 to assemble the required DNA units encoding the catalytic and transport domains into a

DNA assembly that encodes said synthetic enzyme which is capable of synthesising the target molecule.

d) placing the DNA assembly into a vector to allow expression of the synthetic enzyme in a host capable of synthesising the target molecule after transformation with said vector.

28. The method of claim 27 wherein the transformed host is tested for the presence of the target molecule after step d).

29. The transformed host of claim 27.

30. Use of transformed host of claim 27 to produce said target molecule.

31. A method of making a synthetic enzyme to catalyse the synthesis of a target molecule comprising the steps of

a) examining the composition and stereochemistry of a target molecule,

b) determining which catalytic and transport domains need to be present in the synthetic enzyme in order to catalyse the synthesis of the target molecule,

c) using any one of the methods of claims 1-21 to assemble the required DNA units encoding the catalytic and transport domains into a DNA assembly that encodes an enzyme which is capable of synthesising the target molecule.

d) expressing the DNA assembly in a suitable host to produce the enzyme.

32. A library of DNA units encoding catalytic or transport protein domains, wherein each DNA unit has a recognition sequence for a restriction enzyme at it's 5'-end and a second recognition sequence for the same or a compatible enzyme at it's 3'-end which incorporates a recognition sequence for a DNA modifying enzyme.

33. The library of claim 32, wherein each DNA unit has an *Xba*I recognition sequence 5'XXTCTAGA3' (where XX is not GA) at it's 5'-end and an *Xba*I recognition sequence 5'GATCTAGA3' at it's 3'-end

34. A library of DNA units encoding catalytic or transport protein domains, wherein each DNA unit has a recognition sequence at its 5' end for a first restriction enzyme, and a downstream recognition sequence for a second restriction enzyme followed by a downstream recognition sequence for a third restriction enzyme at its 3' end, such that the DNA units, once restricted by the first and second restriction enzymes can be ligated together to abolish the restriction sites at the ligation junction.

35. The library of claim 34, wherein each DNA unit has a *Spe*I recognition sequence 5'ACTAGT3' at its 5'-end, and a downstream *Xba*I recognition sequence 5'TCTAGA3' followed by a downstream *Sma*I recognition sequence 5'CCCGGG3' at it's 3'-end

34. The library of claim 32 or claim 34, wherein the DNA units encode polyketide synthetic domains, comprising two KS domains, at least two AT domains, two KR domains, two DH domains, two ER domains, an ACP domain and a TE domain.

35. A module comprising a DNA sequence encoding a functional set of polyketide synthetic domains wherein the module has a recognition sequence for a restriction enzyme at it's 5'-end and a second recognition sequence for the same or a compatible enzyme at it's 3'-end which incorporates a recognition sequence for a DNA modifying enzyme

36. The module as claimed in claim 35, wherein the module has an *Xba*I recognition sequence 5'XXTCTAGA3' (where XX is not GA) at it's 5'-end and an *Xba*I recognition sequence 5'GATCTAGA3' at it's 3'-end

37. A module comprising a DNA sequence encoding a functional set of polyketide synthetic domains wherein the module has a recognition sequence at its 5' end for a first restriction enzyme, and a downstream recognition sequence for a second restriction enzyme followed a downstream recognition sequence for a third restriction enzyme at its 3' end, such that the DNA units, once restricted by the first and second restriction enzymes can be ligated together to abolish the restriction sites at the ligation junction

38. The module as claimed in claim 37, wherein the module has a *Spe*I recognition sequence 5'ACTAGT3' at its 5'-end, and a downstream *Xba*I recognition sequence 5'TCTAGA3' followed by a downstream *Sma*I recognition sequence 5'CCCGGG3' at it's 3'-end

39. A module as claimed in claim 35 or claim 37, wherein the DNA units encode polyketide synthetic domains, comprising two KS domains, at least two AT domains, two KR domains, two DH domains, two ER domains, an ACP domain and a TE domain

40. A vector containing one or more modules as claimed in claim 35 or claim 37.

41. The vector as claimed in claim 40, wherein a non-functional recA gene is also present.

42. A method of transforming a host with one or more synthetic DNA assemblies encoding enzyme domains which comprises the steps of:

a) Inserting said DNA assembly into a vector containing a mutated internal fragment of a recA gene sequence such that the vector is capable of undergoing homologous recombination with the recA gene of the host,

b) bringing said vector into contact with a host chromosome under conditions which permit homologous recombination to take place,

c) disrupting the host recA gene by the integration of the DNA of said vector into the chromosome.

43. The method of claim 42 wherein the expression vector is used to transform a Steptomyces host.

44. The method of claim 42 or claim 43, wherein the DNA assemblies are modules according to claim 35 or claim 37.

45. A host lacking a recA function, transformed with a vector containing one or more modules according to claim 35 or 37.

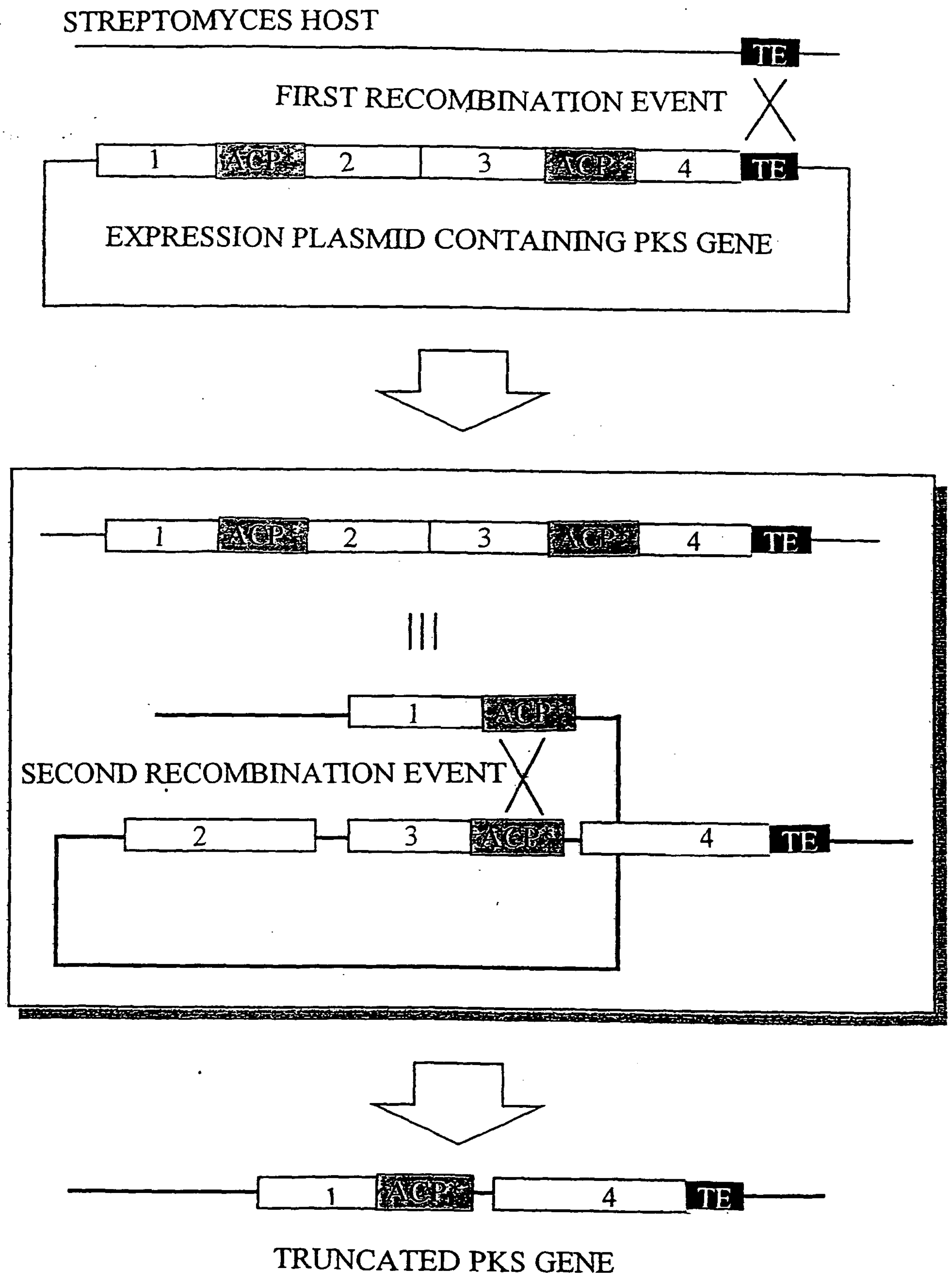
46. A kit containing DNA units, DNA modules, vectors, DNA manipulation hosts, DNA modification hosts, expression hosts, or solid phase elements for use in the methods claimed herein.

KS	AT	KR	DH	ER	ACP
METHYL AS 'UP'	ACETATE	OH AS 'UP'	DOUBLE BOND AS 'E'	METHYL AS 'UP'	ALL HAVE SAME FUNCTION
	PRO-PIONATE				
METHYL AS 'DOWN'	BUTYRATE	OH AS 'DOWN'	DOUBLE BOND AS 'Z'	METHYL AS 'DOWN'	

2	3	2	2	2	1

TOTAL NUMBER OF DOMAINS
REQUIRED FOR EVERY CONCEIVEABLE
POLYKETIDE REACTION= 12

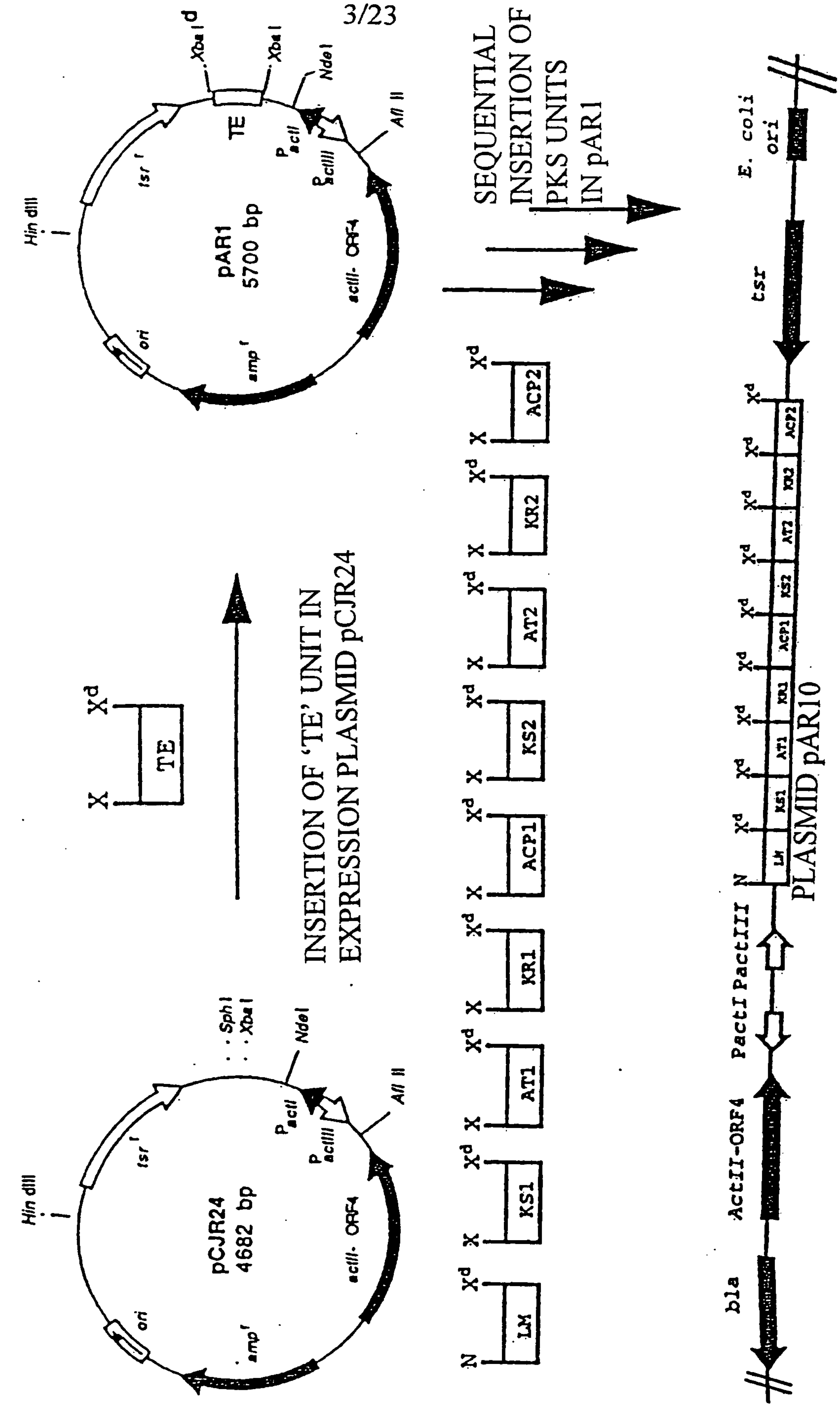
FIG. 1.



STREPTOMYCES HOST
TE
FIRST RECOMBINATION EVENT
1
ACP
2
3
ACP
4
TE
EXPRESSION PLASMID CONTAINING PKS GENE
1
ACP
2
3
ACP
4
TE
1
ACP
SECOND RECOMBINATION EVENT
2
3
ACP
4
TE
1
ACP
4
TE
TRUNCATED PKS GENE

FIG.2.

FIG.3A.


pCJR24
4682 bp
Hin dIII
tsr
Sph I
Xba I
Nde I
Pactl
Pactlll
Afl II
actII- ORF4
amp
ori
pAR1
5700 bp
TE
INSERTION OF 'TE' UNIT IN
EXPRESSION PLASMID pCJR24
SEQUENTIAL
INSERTION OF
PKS UNITS
IN pAR1
N LM X KS1 X AT1 X KR1 X ACP1 X KS2 X AT2 X KR2 X ACP2
bla
ActII-ORF4
PactI PactIII
tsr
E. coli
ori
PLASMID pAR10

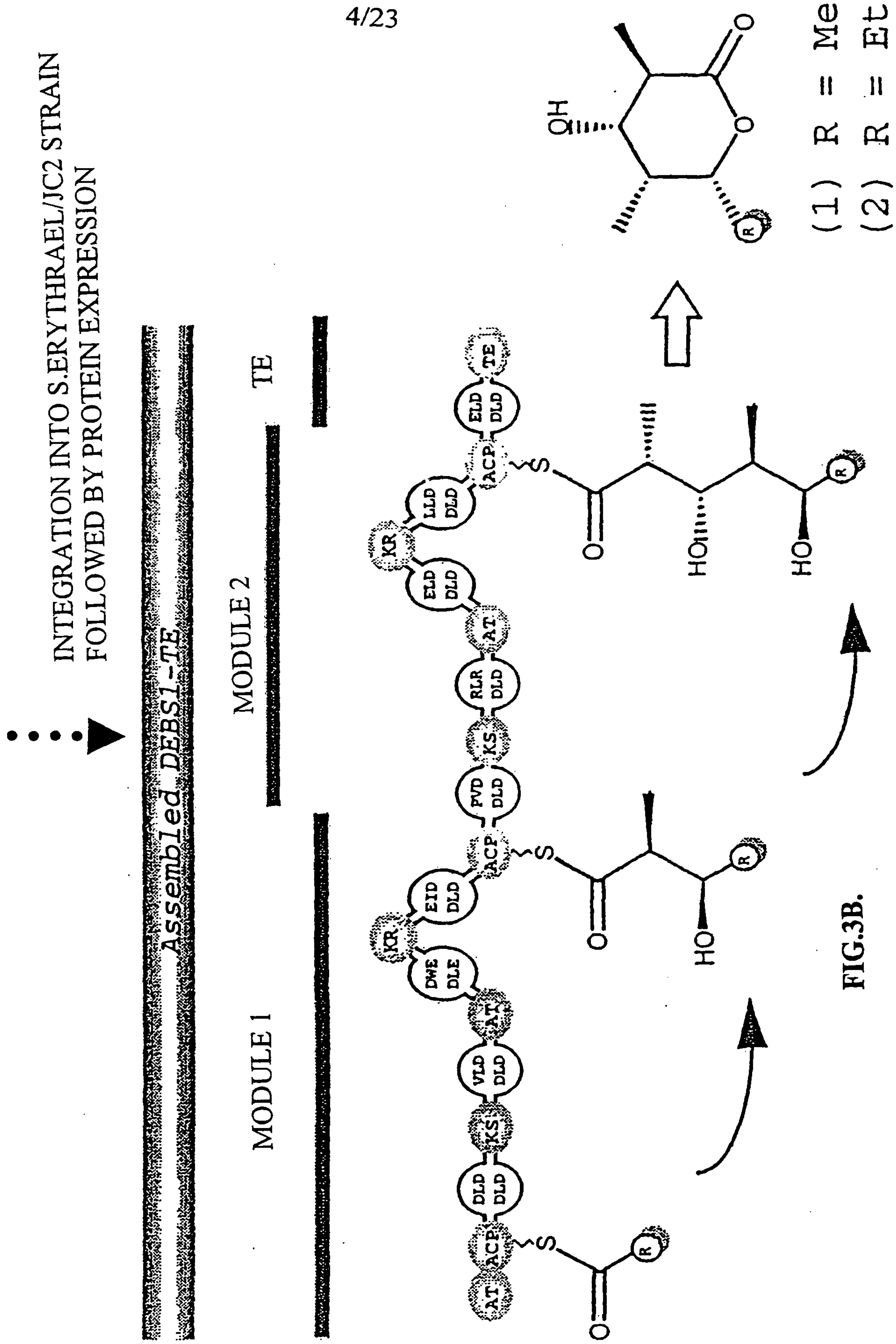

INTEGRATION INTO S.ERYTHRAEL/JC2 STRAIN
FOLLOWED BY PROTEIN EXPRESSION
Assembled DEBS1-TE
MODULE 1
MODULE 2
TE
AT
ACP
DLD DLD
KS
VLD DLD
AT
DWE DLE
KR
EID DLD
ACP
PVD DLD
KS
RLR DLD
AT
ELD DLD
KR
LLD DLD
ACP
ELD DLD
TE
(1) R = Me
(2) R = Et

FIG.3B.

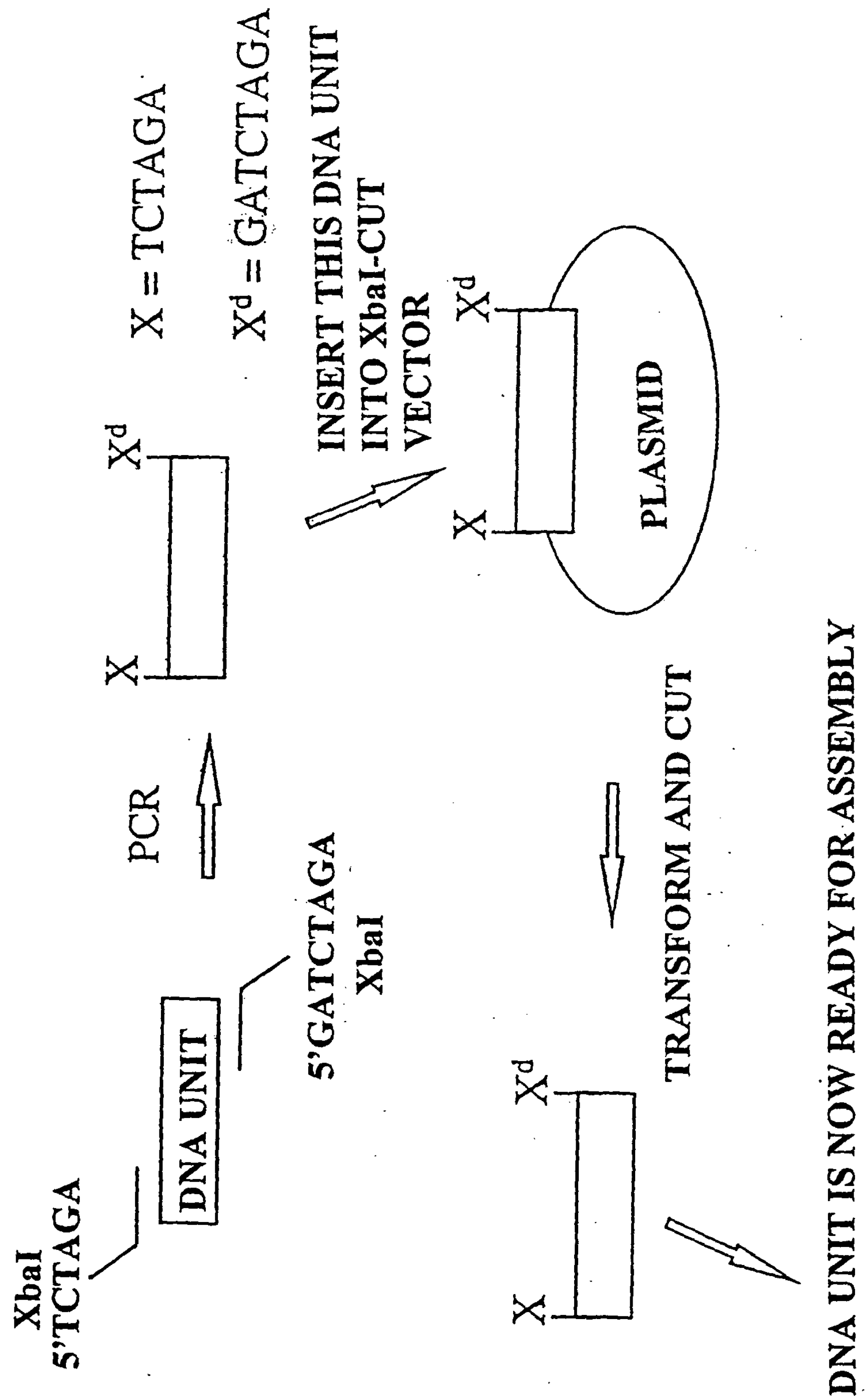

XbaI
5'TCTAGA
DNA UNIT
5'GATCTAGA
XbaI
PCR
X
X^d
X = TCTAGA
X^d = GATCTAGA
INSERT THIS DNA UNIT
INTO XbaI-CUT
VECTOR
X
X^d
PLASMID
TRANSFORM AND CUT
X
X^d
DNA UNIT IS NOW READY FOR ASSEMBLY

FIG.4.

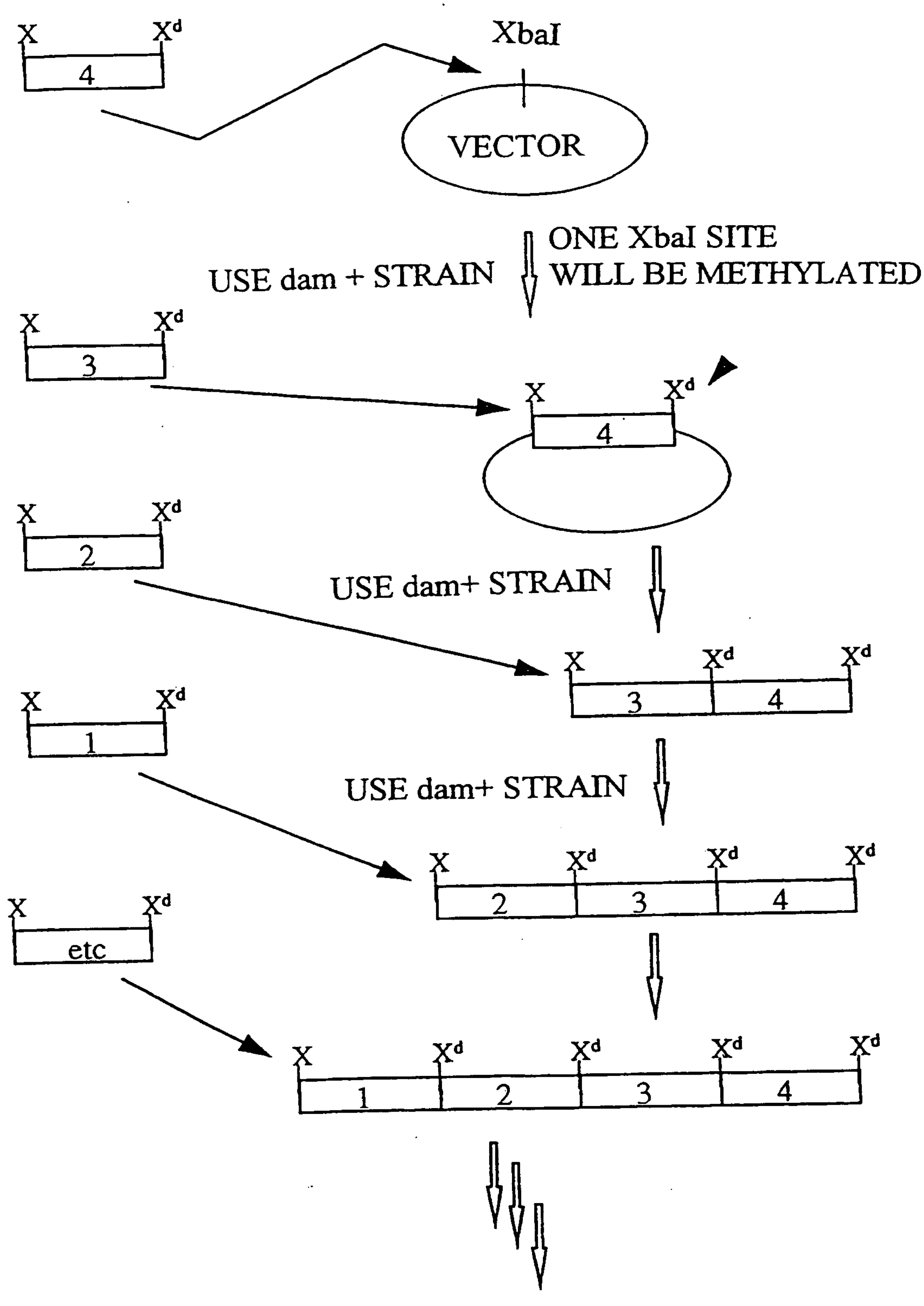

X
Xd
4
XbaI
VECTOR
USE dam + STRAIN
ONE XbaI SITE
WILL BE METHYLATED
3
2
USE dam+ STRAIN
1
etc

FIG.5.

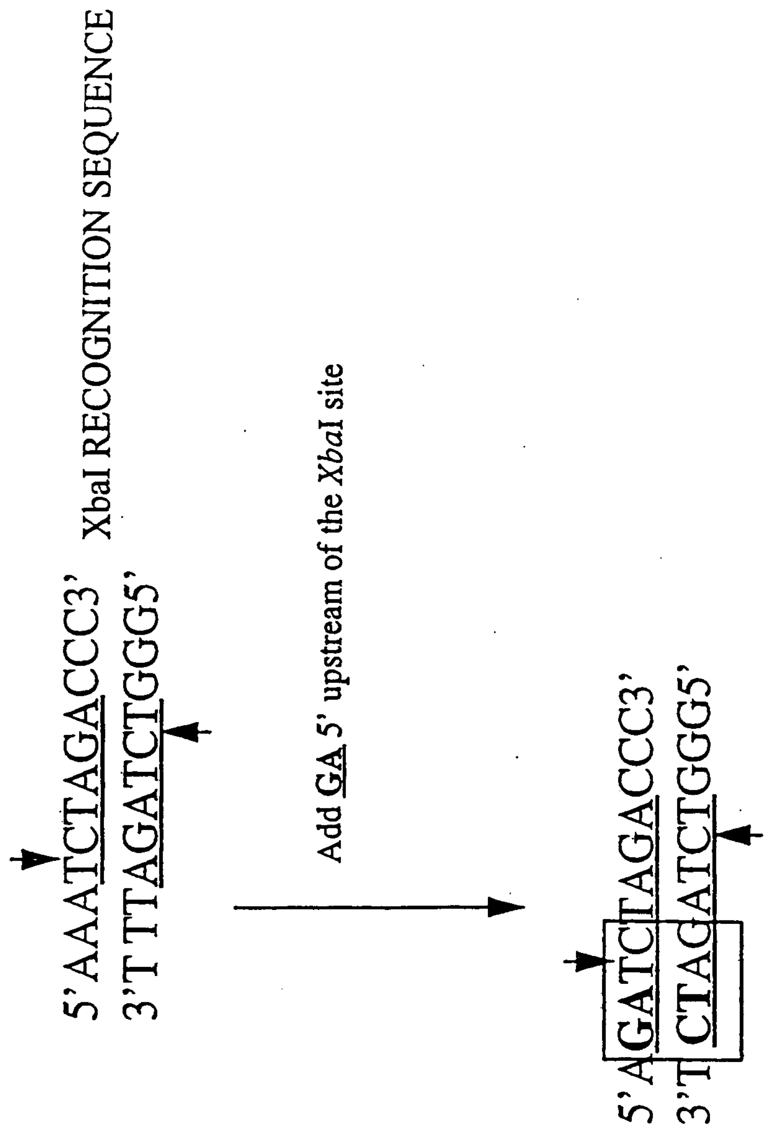

5' AAATCTAGACCC3'
3'T TTAGATCTGGG5'
XbaI RECOGNITION SEQUENCE
Add GA 5' upstream of the XbaI site
5' AGATCTAGACCC3'
3'T CTAGATCTGGG5'

FIG.6.

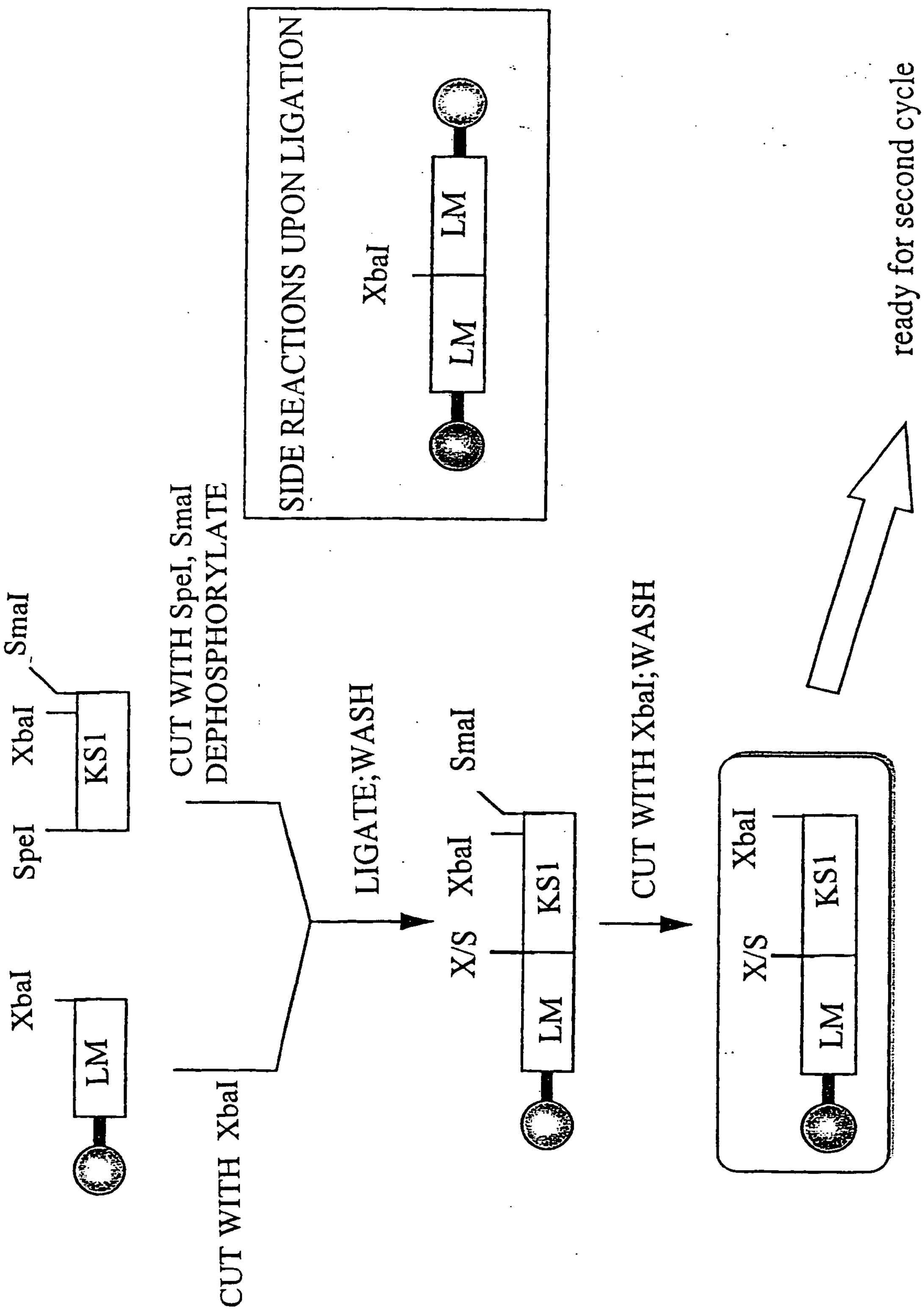

XbaI
LM
SpeI
XbaI
SmaI
KS1
CUT WITH XbaI
CUT WITH SpeI, SmaI
DEPHOSPHORYLATE
LIGATE;WASH
X/S
XbaI
SmaI
LM
KS1
CUT WITH XbaI;WASH
X/S
XbaI
LM
KS1
SIDE REACTIONS UPON LIGATION
XbaI
LM
LM
ready for second cycle

FIG.7.

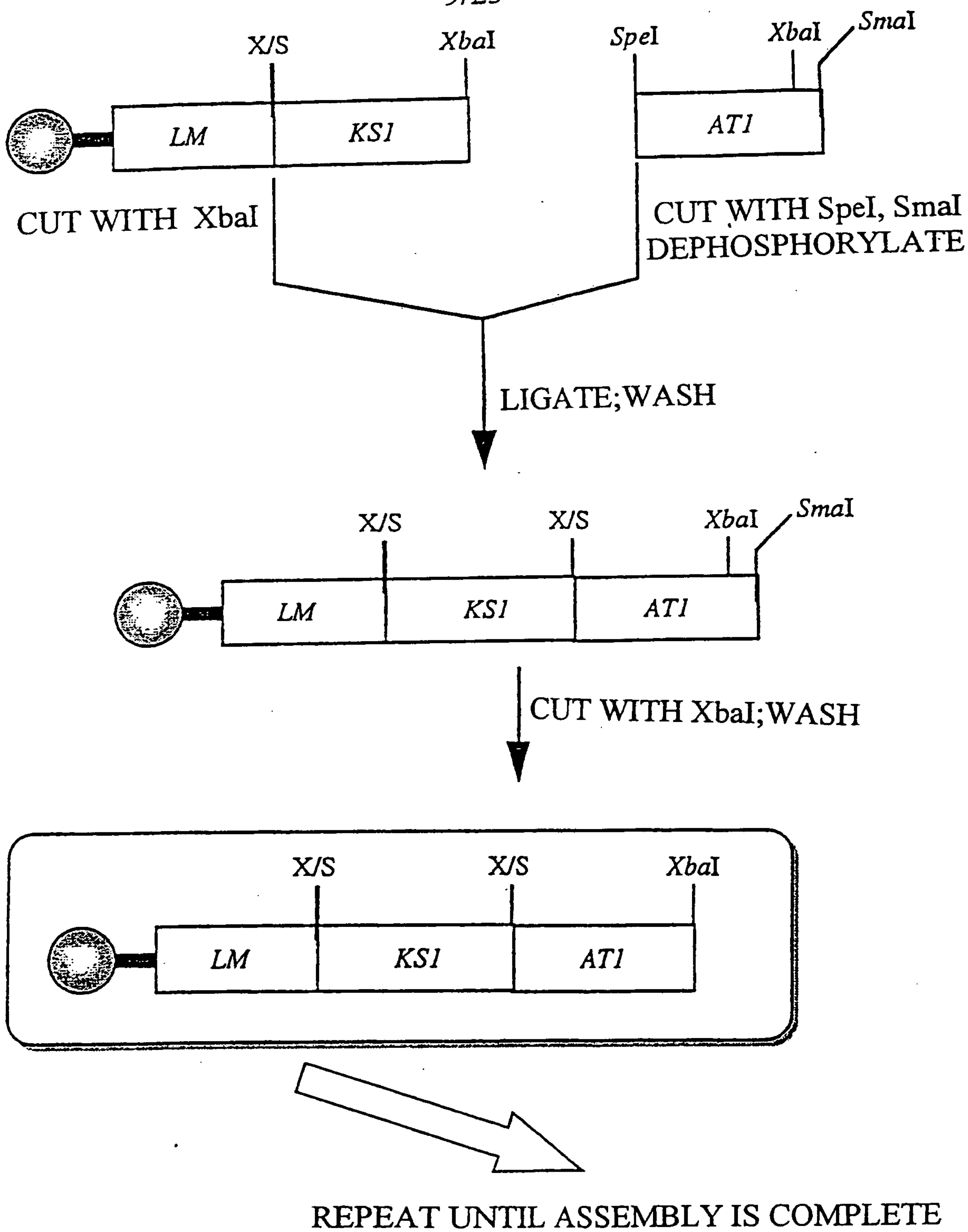
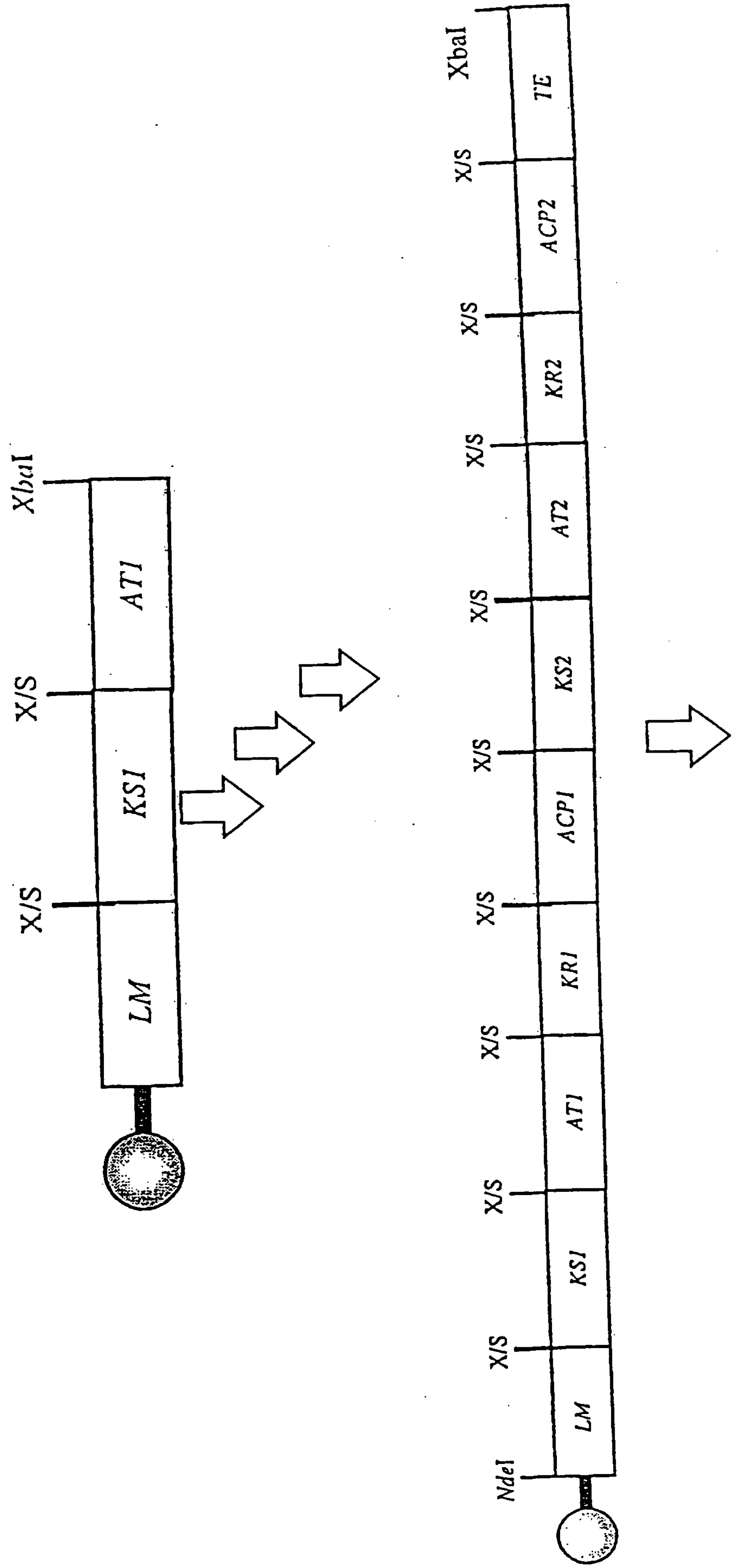

X/S
XbaI
SpeI
XbaI
SmaI
LM
KS1
AT1
CUT WITH XbaI
CUT WITH SpeI, SmaI
DEPHOSPHORYLATE
LIGATE;WASH
X/S
X/S
XbaI
SmaI
LM
KS1
AT1
CUT WITH XbaI;WASH
X/S
X/S
XbaI
LM
KS1
AT1
REPEAT UNTIL ASSEMBLY IS COMPLETE

FIG.8.



CLONE IN pCJR24 AND EXPRESS IN S. ERYTHRAEA

FIG.9.

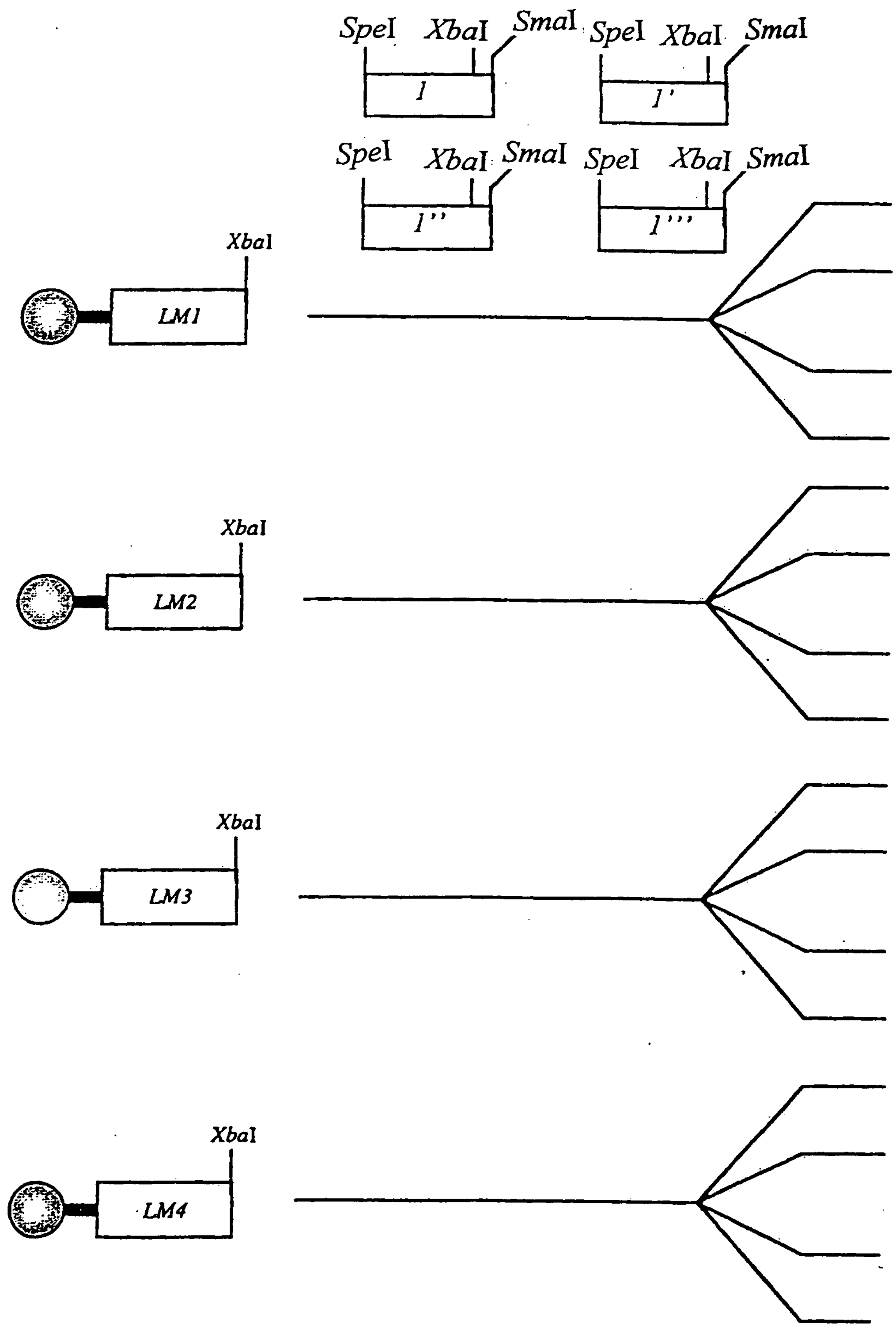

SpeI XbaI SmaI
1
SpeI XbaI SmaI
1'
SpeI XbaI SmaI
1''
SpeI XbaI SmaI
1'''
XbaI
LM1
XbaI
LM2
XbaI
LM3
XbaI
LM4

FIG.10.

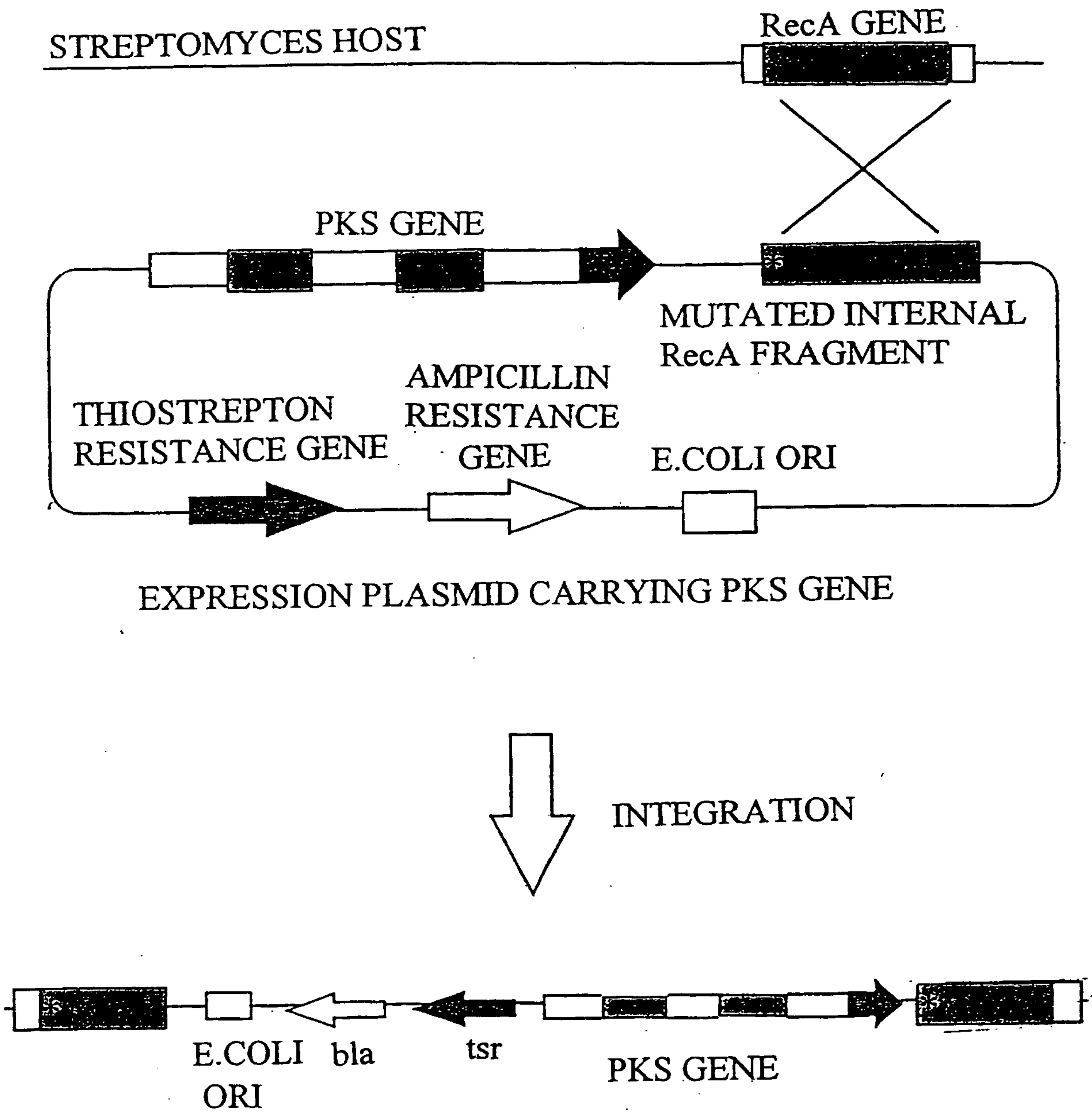

STREPTOMYCES HOST
RecA GENE
PKS GENE
MUTATED INTERNAL
RecA FRAGMENT
AMPICILLIN
RESISTANCE
GENE
THIOSTREPTON
RESISTANCE GENE
E.COLI ORI
EXPRESSION PLASMID CARRYING PKS GENE
INTEGRATION
E.COLI
ORI
bla
tsr
PKS GENE

FIG.11.

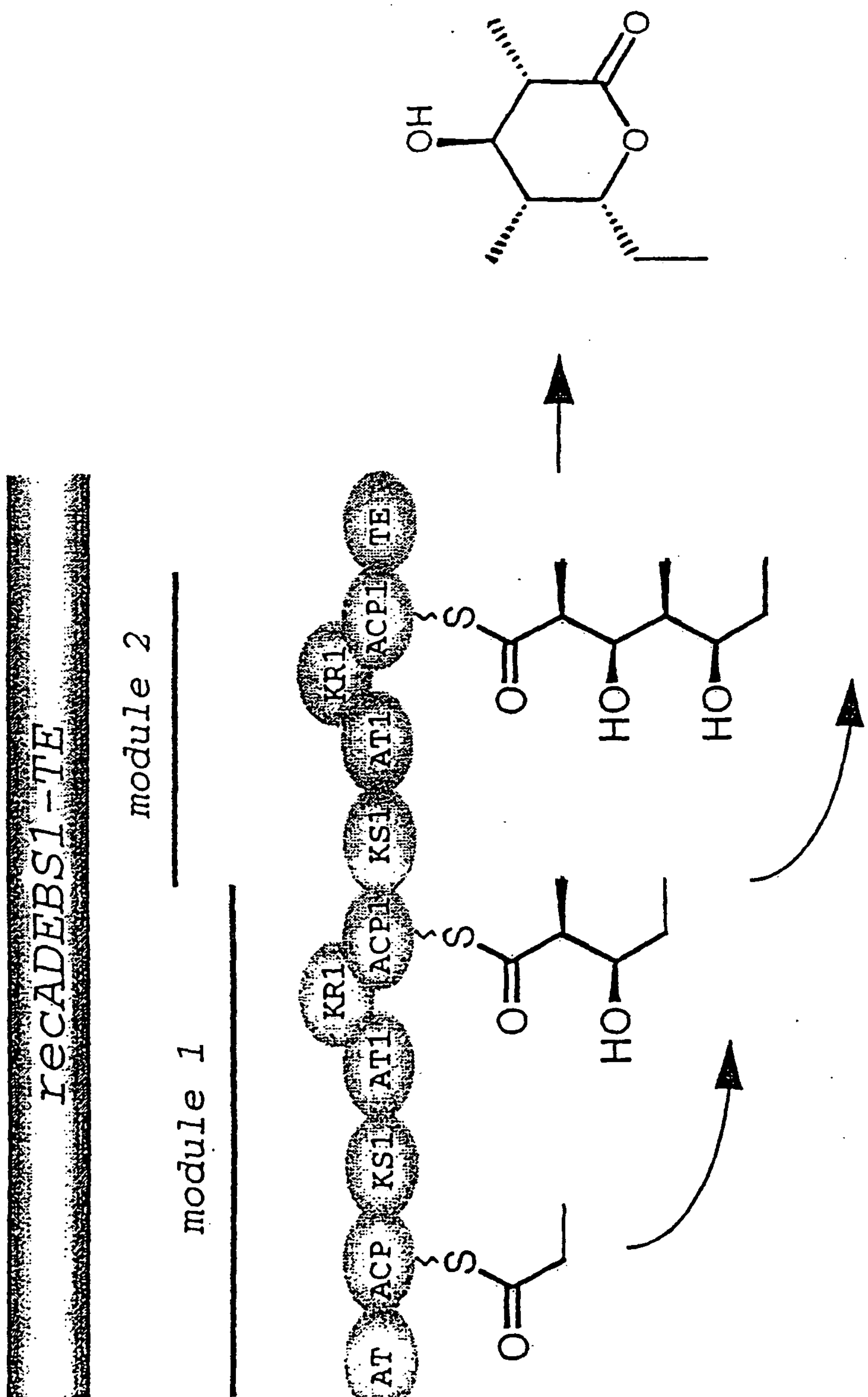


FIG.12.

```
S.l.   1 MAGTDREKALDAALAQIERQFGKGAVMRMGDRTNEPIEVIPTGSTALDVA  50
         ||||||||||||||||||||||||||||||||. ||||||||||||||||
S.a.   1 MAGTDREKALDAALAQIERQFGKGAVMRMGDRSKEPIEVIPTGSTALDVA  50
                  .         .         .         .         .
      51 LGVGGIPRGRVVEVYGPESSGKTTLTLHAVANAQKAGGQVAFVDAEHALD 100
         |||||:|||||:||||||||||||||||||||||||||||||||||||||
      51 LGVGGLPRGRVIEVYGPESSGKTTLTLHAVANAQKAGGQVAFVDAEHALD 100
                  .         .         .         .         .
     101 PEYAKKLGVDIDNLILSQPDNGEQALEIVDMLVRSGALDLIVIDSVAALV 150
         ||||.|||||||||||||||||||||||||||||||||||||||||||||
     101 PEYAQKLGVDIDNLILSQPDNGEQALEIVDMLVRSGALDLIVIDSVAALV 150
                  .         .         .         .         .
     151 PRAEIEGEMGDSHVGLQARLMSQALRKITSALNQSKTTAIFINQLREKIG 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 PRAEIEGEMGDSHVGLQARLMSQALRKITSALNQSKTTAIFINQLREKIG 200
                  .         .         .         .         .
     201 VMFGSPETTTGGRALKFYASVRLDIRRIETLKDGTDAVGNRTRVKVVKNK 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 VMFGSPETTTGGRALKFYASVRLDIRRIETLKDGTDAVGNRTRVKVVKNK 250
                  .         .         .         .         .
     251 VAPPFKQAEFDILYGQGISREGGLIDMGVENGFVRKAGAWYTYEGDQLGQ 300
         ||||||||||||||||||||||||||||||.|||||||||||||||||||
     251 VAPPFKQAEFDILYGQGISREGGLIDMGVEHGFVRKAGAWYTYEGDQLGQ 300
                  .         .         .         .         .
     301 GKENARNFLKDNPDLANEIEKKIKQKLGVGVHPEE.SATEPGADAASAAP 349
         ||||||||||||||||||||||||:|||||| |||  .||| | | |||.
     301 GKENARNFLKDNPDLANEIEKKIKEKLGVGVRPEEPTATESGPDAAT... 347
                  .         .
     350 ADAAPAVPAPTTAKATKSKAAAAKS 374
         |:.||||||| ||| ||.|||||||
     348 AESAPAVPAPATAKVTKAKAAAAKS 372
```

FIG.13.

15/23

```
S.l.    1 ATGGCAGGAACCGACCGCGAGAAGGCCCTGGACGCCGCGCTCGCACAGAT 50
          |||||||||||||||||||||||||| || |||||||| |||||||||||
S.a.    1 ATGGCAGGAACCGACCGCGAGAAGGCTCTTGACGCCGCACTCGCACAGAT 50
                   .         .         .         .         .
       51 TGAACGGCAATTCGGCAAGGGCGCGGTCATGCGCATGGGTGACCGGACCA 100
          ||||||||| ||||||||||||||||||||||||||||| |||||| | |
       51 TGAACGGCAGTTCGGCAAGGGCGCGGTCATGCGCATGGGCGACCGGTCGA 100
                   .         .         .         .         .
      101 ACGAGCCCATCGAGGTCATCCCGACCGGGTCTACCGCGCTCGACGTGGCC 150
          | ||||||||||||||||||||||||||||| ||||||||||||||||||
      101 AGGAGCCCATCGAGGTCATCCCGACCGGGTCGACCGCGCTCGACGTGGCC 150
                   .         .         .         .         .
      151 CTCGGCGTCGGAGGCATCCCGCGTGGCCGTGTCGTGGAGGTCTACGGCCC 200
          ||||||||||| ||| | ||||| ||||| ||| | ||||||||||| ||
      151 CTCGGCGTCGGCGGCCTGCCGCGCGGCCGCGTCATCGAGGTCTACGGTCC 200
                   .         .         .         .         .
      201 CGAGTCCTCGGGCAAGACGACCCTGACCCTGCACGCGGTGGCGAACGCGC 250
           |||||||| || ||||||||||||||||||||||| |||||||||||||
      201 GGAGTCCTCCGGTAAGACGACCCTGACCCTGCACGCCGTGGCGAACGCGC 250
                   .         .         .         .         .
      251 AGAAGGCCGGCGGCCAGGTCGCGTTCGTGGACGCCGAGCACGCCCTCGAC 300
          ||||||||||||||||||| |||||||||||||| |||||||||||||||
      251 AGAAGGCCGGCGGCCAGGTGGCGTTCGTGGACGCGGAGCACGCCCTCGAC 300
                   .         .         .         .         .
      301 CCCGAGTACGCGAAGAAGCTCGGTGTCGACATCGACAACCTGATCCTGTC 350
          |||||||||||  |||||||||| ||||||||||||||||||||||||||
      301 CCCGAGTACGCCCAGAAGCTCGGCGTCGACATCGACAACCTGATCCTGTC 350
                   .         .         .         .         .
      351 CCAGCCGGACAACGGTGAGCAGGCCCTGGAGATCGTGGACATGCTGGTCC 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      351 CCAGCCGGACAACGGTGAGCAGGCCCTGGAGATCGTGGACATGCTGGTCC 400
                   .         .         .         .         .
      401 GCTCCGGCGCCCTCGACCTCATCGTCATCGACTCCGTCGCCGCGCTCGTC 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      401 GCTCCGGCGCCCTCGACCTCATCGTCATCGACTCCGTCGCCGCGCTCGTC 450
                   .         .         .         .         .
      451 CCGCGCGCGGAGATCGAGGGCGAGATGGGCGACAGCCACGTCGGTCTGCA 500
          ||||||||||||||||||||||||||||| ||||||||||||||||| ||
      451 CCGCGCGCGGAGATCGAGGGCGAGATGGGTGACAGCCACGTCGGTCTCCA 500
                   .         .         .         .         .
      501 GGCCCGGCTGATGAGCCAGGCCCTGCGGAAGATCACCAGCGCGCTCAACC 550
          ||||||||||||||||||||| || |||||||||||||||||||||||||
      501 GGCCCGGCTGATGAGCCAGGCGCTCCGGAAGATCACCAGCGCGCTCAACC 550
```

FIG.14.A.

16/23

```
S.l.  551  AGTCCAAGACCACCGCGATCTTCATCAACCAGCTCCGCGAGAAGATCGGC  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||
S.a.  551  AGTCCAAGACCACCGCGATCTTCATCAACCAGCTCCGCGAGAAGATCGGC  600

      601  GTGATGTTCGGCTCCCCGGAGACCACGACCGGTGGCCGGGCACTGAAGTT  650
           || |||||||||||||||||||||||||||||||||||||| || |||||
      601  GTCATGTTCGGCTCCCCGGAGACCACGACCGGTGGCCGGGCGCTCAAGTT  650

      651  CTACGCCTCGGTGCGACTCGACATCCGGCGTATCGAGACGCTGAAGGACG  700
           ||||||||||||||||||||||||||| || ||||||||||| |||||||
      651  CTACGCCTCGGTGCGACTCGACATCCGACGCATCGAGACGCTCAAGGACG  700

      701  GCACCGACGCGGTCGGCAACCGCACCCGCGTCAAGGTGGTCAAGAACAAG  750
           ||||||||||||||||||||||||| ||||||||||| ||||||||||||
      701  GCACCGACGCGGTCGGCAACCGCACGCGCGTCAAGGTCGTCAAGAACAAG  750

      751  GTCGCGCCGCCCTTCAAGCAGGCCGAGTTCGACATCCTCTACGGCCAGGG  800
           ||||||||||||||||||||||||||||||||||||||||||||||||||
      751  GTCGCGCCGCCCTTCAAGCAGGCCGAGTTCGACATCCTCTACGGCCAGGG  800

      801  CATCAGCCGCGAGGGCGGTCTGATCGACATGGGCGTGGAGAACGGCTTCG  850
           |||||||||||||||||| ||||||||||||||||||||| |||||||||
      801  CATCAGCCGCGAGGGCGGCCTGATCGACATGGGCGTGGAGCACGGCTTCG  850

      851  TCCGCAAGGCCGGCGCCTGGTACACGTACGAGGGCGACCAGCTCGGTCAG  900
           |||||||||||||||||||||||||||||||||||||||||||||| |||
      851  TCCGCAAGGCCGGCGCCTGGTACACGTACGAGGGCGACCAGCTCGGCCAG  900

      901  GGCAAGGAGAACGCGCGCAACTTCCTGAAGGACAACCCCGACCTGGCCAA  950
           |||||||||||||||||||||||||||||||||||||||||||| |||||
      901  GGCAAGGAGAACGCGCGCAACTTCCTGAAGGACAACCCCGACCTCGCCAA  950

      951  CGAGATCGAGAAGAAGATCAAGCAGAAGCTGGGCGTCGGCGTGCACCCCG  1000
           ||||||||||||||||||||||| ||||||||||||||| || |  ||||
      951  CGAGATCGAGAAGAAGATCAAGGAGAAGCTGGGCGTCGGAGTCCGTCCCG  1000

     1001  AGGA...GTCGGCCACCGAGCCCGGCGCGGACGCCGCCTCCGCCGCCCCG  1047
           ||||   | ||||||||||| ||||  ||||         ||||||  ||
     1001  AGGAGCCGACGGCCACCGAGTCCGGACCGGA.........CGCCGCGACG  1041

     1048  GCCGACGCCGCACCGGCGGTGCCCGCACCCACGACCGCCAAGGCCACCAA  1097
           |||||  ||||||||||||||||||| ||| |||||||||||| ||||||
     1042  GCCGAATCCGCACCGGCGGTGCCCGCGCCCGCGACCGCCAAGGTCACCAA  1091

     1098  GTCCAAGGCCGCGGCAGCCAAGAGCTGA  1125
           | ||||||||||||||||||||||||||
     1092  GGCCAAGGCCGCGGCAGCCAAGAGCTGA  1119
```

FIG.14.B.

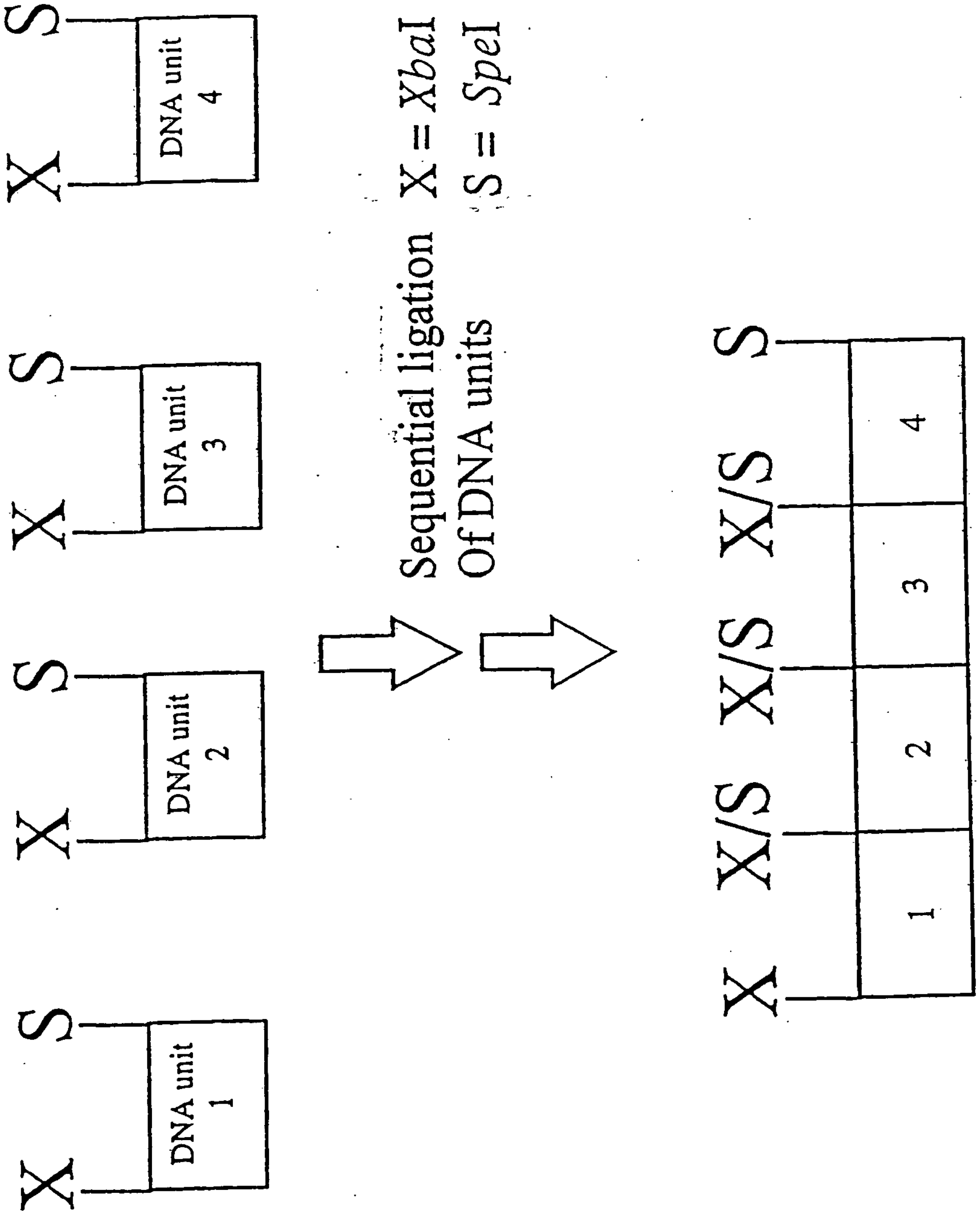
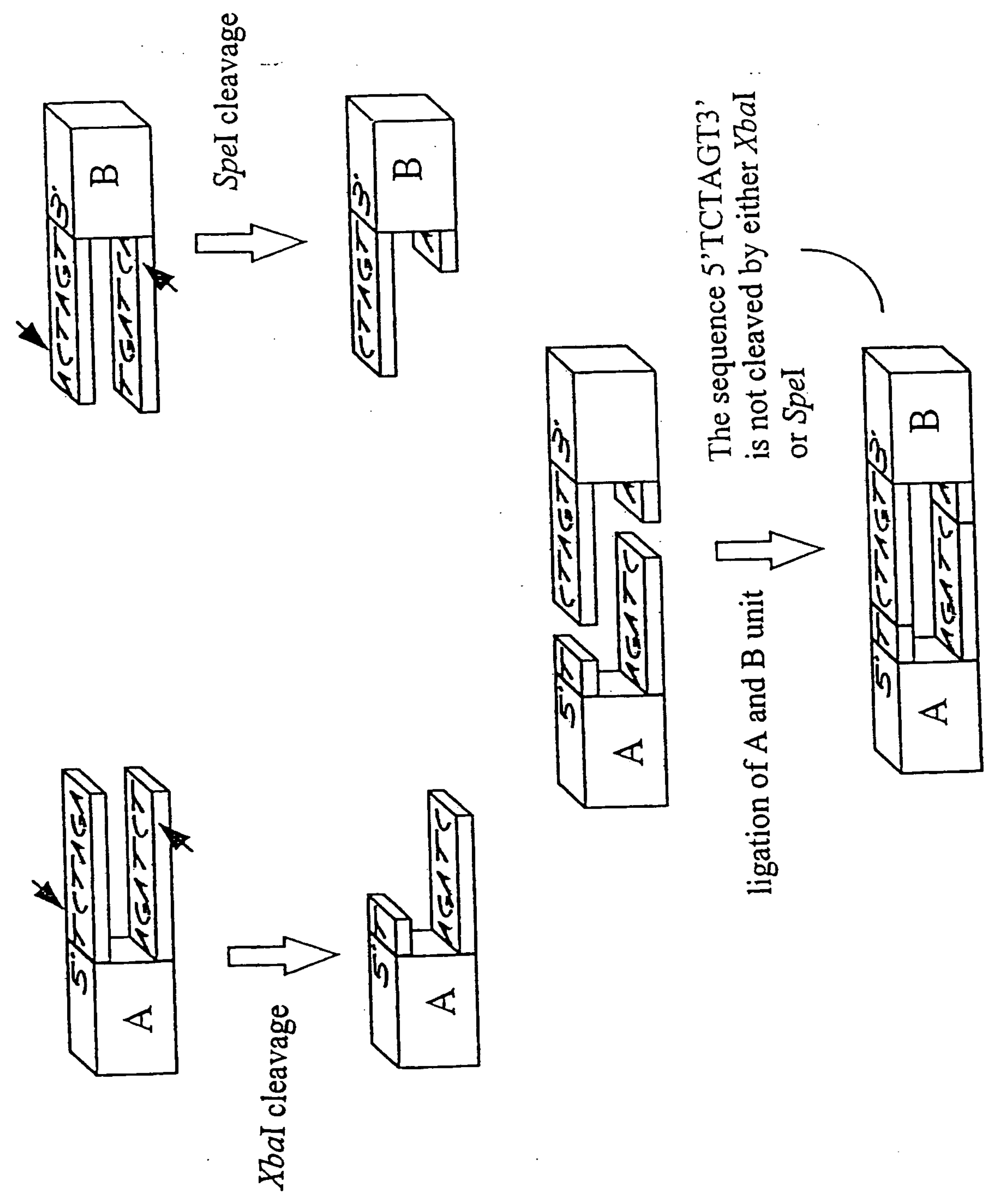

X
S
DNA unit 1
X
S
DNA unit 2
X
S
DNA unit 3
X
S
DNA unit 4
Sequential ligation
Of DNA units
X = XbaI
S = SpeI
X
X/S
X/S
X/S
S
1
2
3
4

Figure 15.



B
SpeI cleavage
B
A
XbaI cleavage
A
A
ligation of A and B unit
The sequence 5'TCTAGT3' is not cleaved by either XbaI or SpeI
A
B

Figure 16.

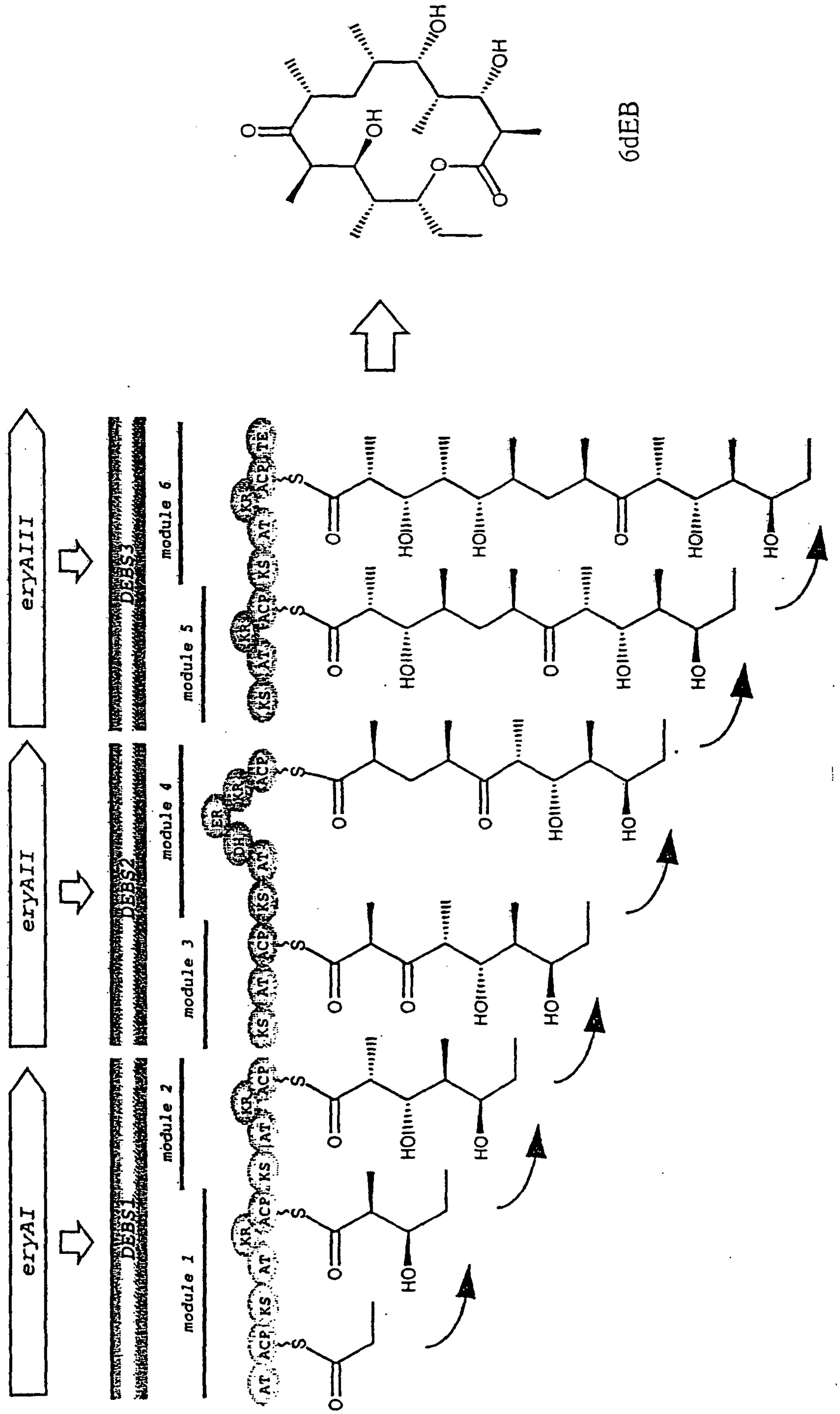

eryAI
eryAII
eryAIII
DEBS1
DEBS2
DEBS3
module 1
module 2
module 3
module 4
module 5
module 6
6dEB

FIG.17.

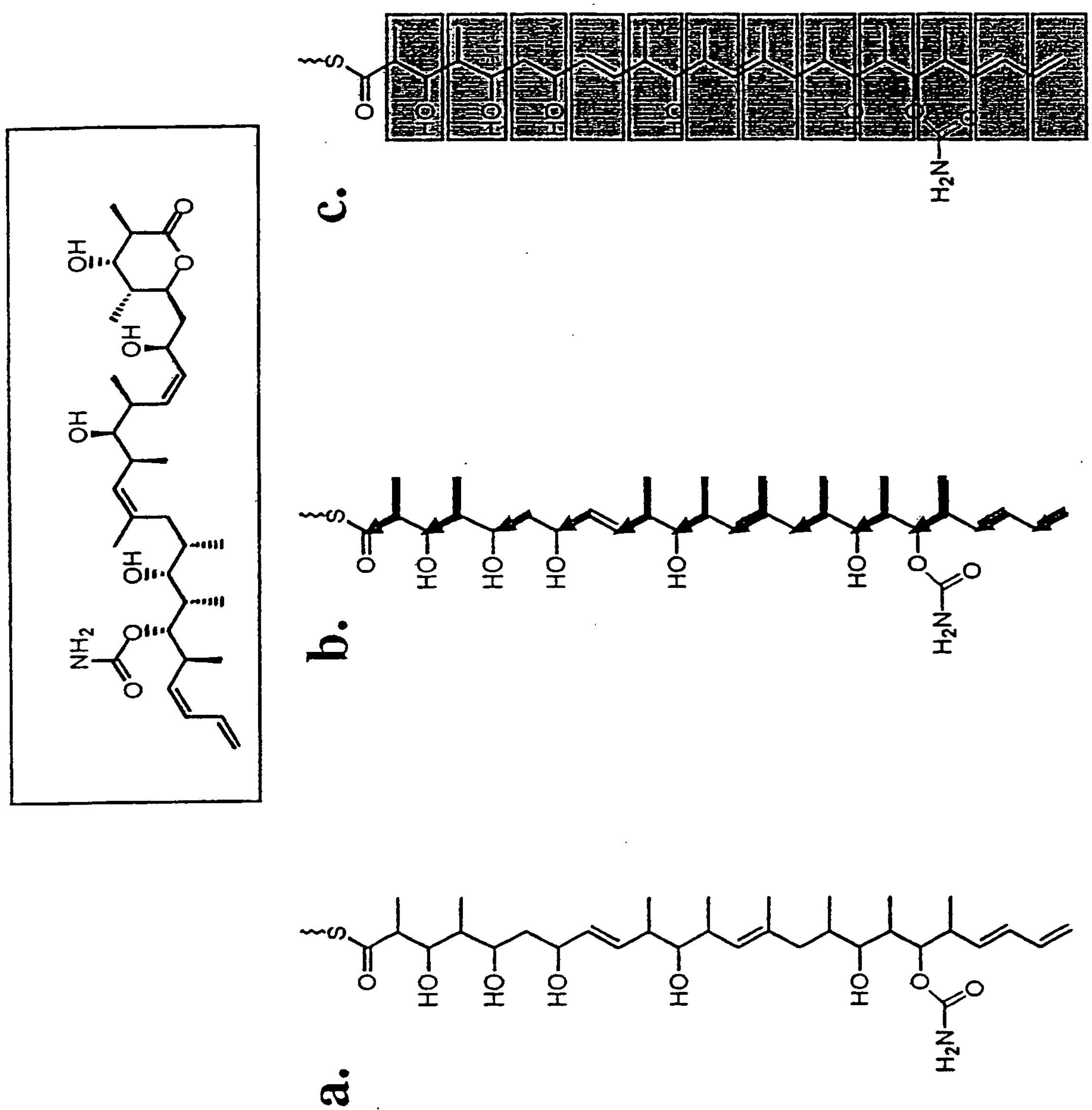


FIG.18.

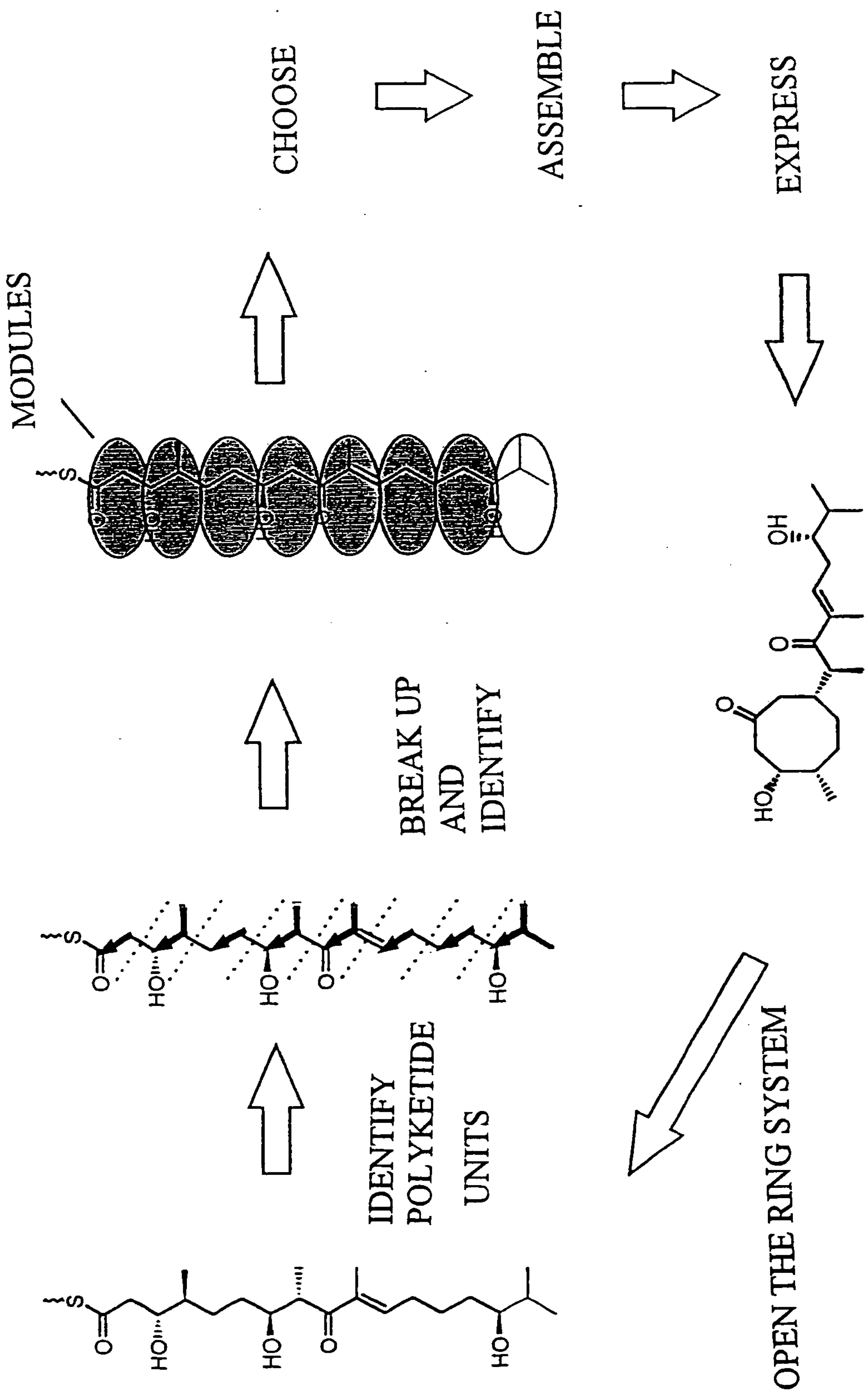

MODULES
CHOOSE
ASSEMBLE
EXPRESS
BREAK UP
AND
IDENTIFY
IDENTIFY
POLYKETIDE
UNITS
OPEN THE RING SYSTEM

FIG.19.

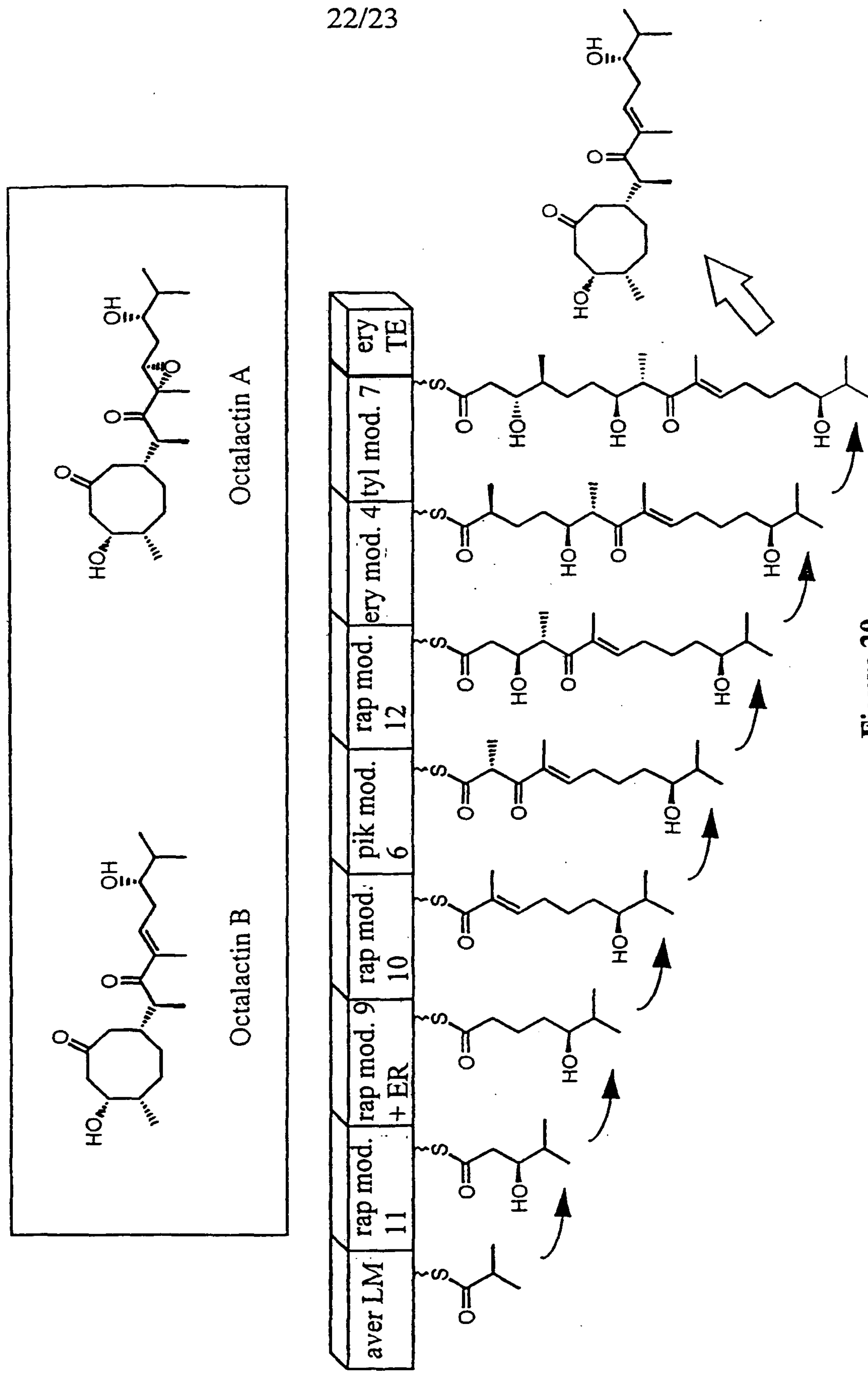

Octalactin A
Octalactin B
aver LM
rap mod. 11
rap mod. 9 + ER
rap mod. 10
pik mod. 6
rap mod. 12
ery mod. 4
tyl mod. 7
ery TE

Figure 20.

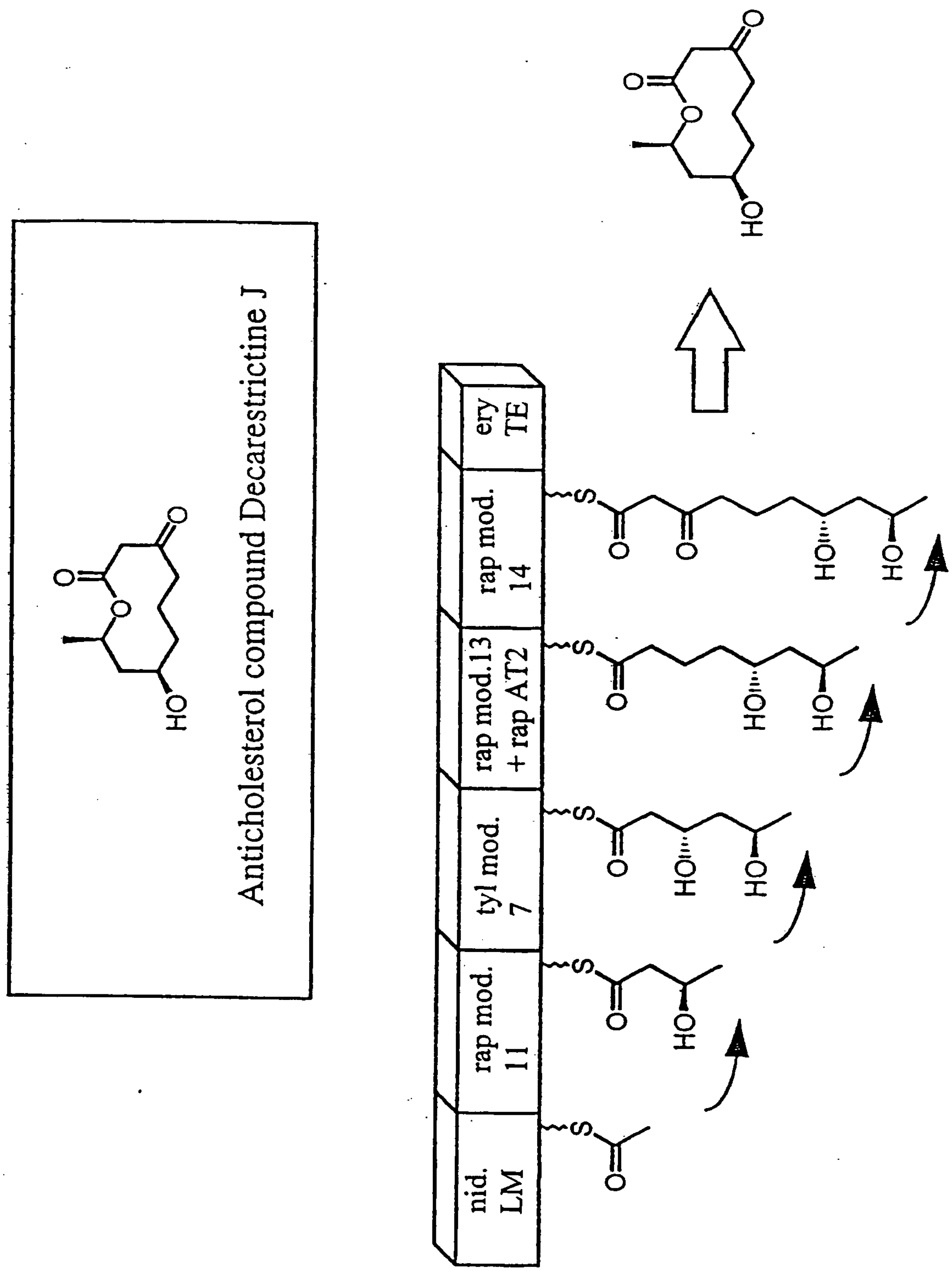

Anticholesterol compound Decarestrictine J
nid. LM
rap mod. 11
tyl mod. 7
rap mod.13 + rap AT2
rap mod. 14
ery TE

Figure 21.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 December 2000 (21.12.2000)

PCT

(10) International Publication Number
WO 00/77181 A3

(51) International Patent Classification[7]: C12N 15/10, 15/66, 15/52, 15/90, C12P 17/06, 17/08

(21) International Application Number: PCT/GB00/02286

(22) International Filing Date: 12 June 2000 (12.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9913694.7 11 June 1999 (11.06.1999) GB

(71) Applicant (for all designated States except US): QXYZ LIMITED [GB/GB]; Mitchells Roberton, Solicitors, George House, 36 North Hanover Street, Glasgow G1 2AD (GB).

(72) Inventor; and
(75) Inventor/Applicant (for US only): RANGANATHAN, Anand [IN/IN]; International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, P.O. Box 10504, New Delhi 110067, Maharashtra (IN).

(74) Agent: HEATON, Joanne, Marie; Stevens, Hewlett & Perkins, 1 St Augustine's Place, Bristol BS1 4UD (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

(88) Date of publication of the international search report: 10 May 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/77181 A3

(54) Title: DNA MANIPULATION METHODS, APPLICATIONS FOR SYNTHETIC ENZYMES AND USE FOR POLYKETIDE PRODUCTION

(57) Abstract: The invention comprises a method of assembling several DNA units in sequence in a DNA construct and all derivatives of this method. In particular the production of synthetic enzymes is contemplated. Each DNA unit is provided with the same restriction enzyme recognition site at its 5' and 3' ends. The restriction recognition site at its 3' end being combined with a recognition site for a DNA modification enzyme. A DNA construct having the same or a compatible accessible restriction site, as provided in the DNA unit, is cleaved at the restriction site by the appropriate restriction enzyme. The desired DNA unit is then inserted into the DNA construct, this ligated product subsequently being brought into contact with a DNA modification enzyme such that the restriction site at the 3' end of the inserted DNA unit is abolished. The ligated product is then cleaved at the remaining unmodified restriction recognition site and a subsequent DNA unit is inserted. This process is repeated introducing each desired DNA unit to give a DNA construct containing all the desired units in sequence.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/02286

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12N15/66 C12N15/52 C12N15/90 C12P17/06
C12P17/08

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, PAJ, CAB Data, STRAND, EPO-Internal, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 17811 A (CHROMAXOME CORP) 30 April 1998 (1998-04-30) claims 1-26; figure 5E; example 5.5.5.	22-26
X	ROWE C J ET AL: "Construction of new vectors for high-level expression in actinomycetes" GENE,NL,ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, vol. 216, no. 1, August 1998 (1998-08), pages 215-223, XP004149299 ISSN: 0378-1119 cited in the application the whole document	22-26

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 February 2001	27.02.01

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hornig, H

3

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 49315 A (KOSAN BIOSCIENCES INC ;UNIV LELAND STANFORD JUNIOR (US)) 5 November 1998 (1998-11-05) claims 1-24; figure 6A	22-26
X	WO 96 40968 A (UNIV LELAND STANFORD JUNIOR ;JOHN INNES CENTRE (GB)) 19 December 1996 (1996-12-19) the whole document	22-26
X	MCDANIEL R ET AL: "Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel unnatural natural products" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 96, no. 5, March 1999 (1999-03), pages 1846-1851, XP002143433 ISSN: 0027-8424 the whole document	22-26
X	MUTH G ET AL: "Mutational analysis of Streptomyces lividans recA gene suggests that only mutants with residual activity remain viable." MOLECULAR & GENERAL GENETICS, vol. 255, no. 4, 1997, pages 420-428, XP002160032 ISSN: 0026-8925 the whole document	44,45
X	EP 0 841 402 A (NAT INST AGROBIO RES) 13 May 1998 (1998-05-13) page 3, line 28 - line 33; claim 19	44
X	US 4 963 487 A (SCHIMMEL PAUL R) 16 October 1990 (1990-10-16) column 2, line 9 - line 55	44
X	US 4 713 337 A (JASIN MARIA ET AL) 15 December 1987 (1987-12-15) column 4, line 66 -column 6, line 5	44
A	US 5 863 730 A (MASSON ET AL.) 26 January 1999 (1999-01-26) claims 1-17	
T	WO 00 63360 A (CELLTECH THERAPEUTICS LTD ;FINNEY HELENE MARGARET (GB); LAWSON ALA) 26 October 2000 (2000-10-26) the whole document	

-/--

3

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 98 38326 A (ZINK MARY ANN ;XU GUOPING (US); HODGSON CLAGUE P (US); NATURE TECH) 3 September 1998 (1998-09-03) the whole document ---	
A	WO 97 28282 A (STRATAGENE INC) 7 August 1997 (1997-08-07) the whole document ---	
A	TER HAAR ERNST ET AL: "Discodermolide, a cytotoxic marine agent that stabilizes microtubules more potently than taxol." BIOCHEMISTRY, vol. 35, no. 1, 1996, pages 243-250, XP002154629 ISSN: 0006-2960 cited in the application the whole document ---	
A	NERENBERG J B ET AL: "TOTAL SYNTHESIS OF THE IMMUNOSUPPRESSIVE AGENT (-)-DISCODERMOLIDE" JOURNAL OF THE AMERICAN CHEMICAL SOCIETY,US,AMERICAN CHEMICAL SOCIETY, WASHINGTON, DC, vol. 115, no. 26, 1993, pages 12621-12622, XP000652058 ISSN: 0002-7863 the whole document ---	
A	TAPIOLAS D M ET AL: "OCTALACTINS A AND B CYTOTOXIC EIGHT-MEMBERED-RING LACTONES FROM A MARINE BACTERIUM STREPTOMYCES-SP" JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 113, no. 12, 1991, pages 4682-4683, XP002154630 ISSN: 0002-7863 cited in the application the whole document ---	
A	YAMADA SHINYA ET AL: "Total synthesis of (-)-decarestrictine J." BIOSCIENCE BIOTECHNOLOGY AND BIOCHEMISTRY, vol. 59, no. 9, 1995, pages 1657-1660, XP002154631 ISSN: 0916-8451 cited in the application the whole document ---	

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,A	RANGANATHAN ANAND ET AL: "Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: A route to simple statin analogues." CHEMISTRY & BIOLOGY (LONDON), vol. 6, no. 10, October 1999 (1999-10), pages 731-741, XP000971117 ISSN: 1074-5521 the whole document	

3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 00/02286

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: (1-8,32,33,37,38)-complete, (16-31,36,41-43, 46-48)-partially

A method of assembling several DNA units in sequence in a DNA construct, which method comprises the step of: a) providing each DNA unit with a restriction enzyme recognition sequence at its 5' end and with a recognition sequence for the same restriction enzyme at its 3' end that is combined with a restriction site for a DNA modification enzyme, b) providing a starting DNA construct having an accessible restriction site for the same or a compatible restriction enzyme and cleaving the starting DNA construct with a restriction enzyme, c) inserting the desired DNA unit and bringing the ligated product into contact with a DNA modification enzyme such that the restriction site at the 3' end of the inserted DNA unit is abolished, d) cleaving the ligated product at an accessible unmodified recognition site for the same or a compatible restriction enzyme, e) repeating step c) and d) to introduce each desired DNA unit to give a DNA construct containing all the desired units in sequence; DNA construct incorporating one or more DNA assemblies encoding synthetic enzymes and/or hosts expressing DNA constructs made by said method; compounds produced by synthetic enzymes encoded by said DNA assemblies; a method of synthesising a target molecule using said method; a method of making a synthetic enzyme to catalyse the synthesis of a target molecule using said method; a library of DNA units encoding a catalytic or transport protein domains, wherein each DNA unit has a recognition sequence for a restriction enzyme at its 5'-end and a second recognition sequence for the same or a compatible enzyme at its 3'-end which incorporates a recognition sequence for a DNA modifying enzyme;
a module comprising a DNA sequence encoding a functional set of polyketide synthetic domains wherein the module has a recognition sequence for a restriction enzyme at its 5'-end and a second recognition sequence for the same or a compatible enzyme at its 3'-end which incorporates a recognition sequence for a DNA modifying enzyme;
a method of transforming a host with one or more synthetic DNA assemblies encoding enzyme domains, wherein the DNA assemblies are said modules;

2. Claims: (9-15,33,34,39,40)-complete, (16-31,36,41-43, 46-48)-partially

Idem as invention 1, but limited to a method of: assembling several DNA units in sequence in a DNA construct, which method comprises the step of: a) providing a first DNA unit

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

with a recognition sequence for a first restriction enzyme at its 3' end, and cleaving the said first DNA unit with said first restriction enzyme, b) providing each other DNA unit with a recognition sequence at its 5' end for a second restriction enzyme which has a compatible ligation sequence with that of the first restriction enzyme, and a downstream recognition sequence for said first restriction enzyme followed by a downstream recognition sequence for a third restriction enzyme at its 3' end, and cleaving each said other DNA unit with the second and third restriction enzymes, c) ligating the said first DNA unit with a desired other DNA unit to form a ligated product such that the ligation of the two units abolishes the recognition site for the first restriction enzyme at the ligation junction, and cleaving the ligated product with said first restriction enzymel, d) ligating the product from c) with a desired DNA unit from b) to form a ligated product and cleaving the ligated product with said first restriction enzyme, e)repeating step d) with each other DNA unit in turn so as to assemble the DNA unit in sequence;

3. Claims: 44-45

A method of transforming a host with one or more synthetic DNA assemblies encoding enzyme domains which comprises the step of: a) Inserting said DNA assembly into a vector containing a mutated internal fragment of a recA gene sequence such that the vector is capable of undergoing homologous recombination with the recA gene of the host, b) bringing said vector into contact with a host chromosome under conditions which permit homologous recombination to take place, c) disrupting the host recA gene by integration of the DNA of said vector into the chromosome; said method wherein the expression vector is used to transform a Streptomyces host;

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 00/02286

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9817811 A	30-04-1998	US 5783431 A	21-07-1998
		AU 5163298 A	15-05-1998
		EP 0951557 A	27-10-1999
WO 9849315 A	05-11-1998	AU 7172298 A	24-11-1998
		EP 0979286 A	16-02-2000
		US 6117659 A	12-09-2000
WO 9640968 A	19-12-1996	US 5712146 A	27-01-1998
		AU 703920 B	01-04-1999
		AU 6157596 A	30-12-1996
		CA 2224104 A	19-12-1996
		EP 0871760 A	21-10-1998
		NZ 310729 A	29-09-1999
		US 6077696 A	20-06-2000
		US 5962290 A	05-10-1999
EP 0841402 A	13-05-1998	AU 694393 B	16-07-1998
		AU 3922697 A	02-04-1998
		AU 721577 B	06-07-2000
		AU 6804898 A	30-07-1998
		CA 2216596 A	26-03-1998
		CN 1182796 A	27-05-1998
		JP 10155485 A	16-06-1998
		US 6165780 A	26-12-2000
US 4963487 A	16-10-1990	US 4713337 A	15-12-1987
		US 4774180 A	27-09-1988
		EP 0257095 A	02-03-1988
		JP 63502723 T	13-10-1988
		WO 8705331 A	11-09-1987
US 4713337 A	15-12-1987	US 4963487 A	16-10-1990
US 5863730 A	26-01-1999	FR 2738841 A	21-03-1997
		EP 0771872 A	07-05-1997
WO 0063360 A	26-10-2000	NONE	
WO 9838326 A	03-09-1998	AU 6443298 A	18-09-1998
WO 9728282 A	07-08-1997	NONE	

REPLACED BY ART 34 AMDT

Identify polyketide units – e.g. whether acetate, propionate, etc,

Break-up and identify – break up the carbon skeleton and identify how many such carbon units are present. Eight units would mean one requires eight modules to make a PKS.

Choose – choose the modules or domains that would be required, form an existing library of such PKS modules and domains.

Assemble – assemble the DNA units (modules/domains/using the invention.

Express – express the assembled gene in a host and check for compound production.

Figure 20 shows a schematic representation of they hypothetical polyketide synthase for synthesising octalactin B, assembled from enzyme units that belong to various PKSs in the public domain.

Figure 21 shows a schematic representation of the hypothetical decarestrictine polyketide synthase for synthesising the anti-cholesterol compound decarestrictine J, assembled from enzyme units that belong to various PKSs in the public domain.

Examples

Example 1: Vectorial assembly of DNA units

DNA units that are to be assembled contain the *Xba*I recognition sequence at either end of the unit. At one of the ends, two nucleotides (GA) are arranged at the 5' end of the *Xba*I recognition sequence (thus making it 5'**GA**TCTAGA3'). This is achieved by first incorporating the *Xba*I recognition sequences in the oligonucleotide primers and then amplifying the desired DNA unit by PCR. The PCR products are then ligated to a pUC-18 vector, used to transform a dam$^+$ strain of *E. coli,* and the clones isolated and sequenced for possible errors in the PCR products. A dam$^+$ strain of *E. coli* - like DH10B™ - methylate the nucleotide **A** in the sequence GATCTAGA, as 5'GATC3' is a sequence that is recognised by the product of the Dam methylase gene (Fujimoto *et al.*,1965; Geier *et al.*, 1979). This makes only one end of the DNA unit cleavable by *Xba*I. The vector is then used to transform a dam$^-$ strain of *E. coli* (e.g. ET12567 - MacNeil *et al.* (1992)) and the plasmid DNA isolated. This DNA is now cleavable at *both* ends of the DNA unit by *Xba*I. When a library of units has been constructed using this strategy, and both ends of these units have been cleaved by *Xba*I, they are progressively inserted into a vector that has a unique *Xba*I site and the ligated products are used always to transform a dam$^+$ strain of *E. coli*, thereby making sure that one end of the DNA unit is always protected from cleavage by *Xba*I through methylation. When the assembly of such units is completed, the final plasmid is integrated into a streptomyces strain for the production of the desired polyketide.

Using this methodology, the polyketide synthase DEBS1-TE, a multienzyme that has the first of the three bimodular erythromycin DEBS enzymes (DEBS1), fused with the erythromycin thioesterase (Cortés *et al.*, 1995) was constructed in a *de novo* fashion. The ten inherent PKS domains in DEBS1-TE, namely, loading module (itself composed of an AT and an ACP), KS1 (ketosynthase of module 1), AT1, KR1, ACP1, KS2

(ketosynthase of module 2), AT2, KR2, ACP2 and TE function in conjunction to catalyse the synthesis of (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-hexanoic acid δ-lactone (2), figure 3.

The DNA for all ten domains was amplified by PCR to incorporate the two aforementioned recognition sequences for *Xba*I (5'TCTAGA3' and 5'GATCTAGA3') at the 5' and 3' ends of the DNA unit respectively. The PCR products were cloned in pUC18 vector, sequenced, and then used to transform the dam$^-$ *E. coli* ET12567 strain. To initiate the assembly process, the DNA unit for TE was inserted into *S. erythraea* expression vector pCJR24 (Rowe *et al.*, 1998) which has a unique *Xba*I site. This vector also contains a thiostrepton-resistance gene as a marker for identifying successful integrands. The ligated products were used to transform the dam$^+$ *E. coli* DH10B™ strain and the plasmid DNA isolated. This plasmid (pAR1) can only be singly cleaved with *Xba*I, despite possessing two *Xba*I recognition sequences, as one of the sites (situated at the 3' end of the TE unit) has been methylated by the *E. coli* Dam methylase. The next DNA unit (ACP2 from module 2 of DEBS1) was then ligated to the *Xba*I-cut pAR1, the ligation mixture used to transform DH10B cells and the plasmid DNA isolated. Likewise, the other eight DNA units were successively added to pAR1 to finally yield the expression plasmid pAR10 containing the reconstituted DEBS1-TE gene (Figure 3). The junctions where these domains were joined were chosen in the linker regions that lie between these domains, so as to cause minimum disturbance of the structural features of these domains, that might in tum affect the proficiency of the domains themselves (Figure 3). Plasmid pAR10 was then used to transform *S. erythraea/JC2* - a mutant strain of the wild-type *S. erythraea NRRL2338* that lacks the DEBS genes except for the TE DNA fragment (Rowe *et al.*, 1998). Thiostrepton-resistant colonies were selected upon integration of the vector into the *S. erythraea* chromosome. Single transformants were grown on selective media, as described in the *methods* section. The fermentation broth was extracted with ethyl acetate

and a sample of the organic extract was analysed by gas chromatography-mass spectroscopy (GC-MS). Two peaks were observed, corresponding to molecular massess 158 and 172, indicating the presence of the expected acetate- and propionate- derived polyketides (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-pentanoic acid d-lactone (**1**) and (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-hexanoic acid d-lactone (**2**). Both compounds were isolated and fully characterised by high-pressure liquid chromatography (HPLC), ^{1}H 1D and 2D NMR, ^{13}C NMR, FT-ICR spectrometry, and by comparison with a synthetic standard of (**2**) (Brown *et al.*, 1995). One litre of fermentation broth produces 24 mg of (**1**) and 56 mg of (**2**) - yields that are comparable to those reported elsewhere (Lau *et al.*, 1999). It can therefore be asserted that the ten newly constructed inter-domain junctions have not in any way dimmed the catalytic proficiency of the DEBS1-TE synthase.

In the absence of any crystal-structure data on PKS domains, all genetic engineering efforts known in the art have been based on trial-and-error methods of experimenting with where to join two such domains. As a result, the yield of the synthesised polyketide products have varied depending upon the position in the polypeptide chain at which the domains or modules have been linked (McDaniel *et al.*, 1999; Ruan *et al.*, 1997). The successful functioning of the reconstructed polyketide synthase described above has supplied new information about the inter-domain junction sites. Using this information, and the described methodology for the rapid assembly of these enzyme units, it is now possible to carry out a 'retrobiosynthetic analysis' of target molecules and then to use polyketide and other biosynthetic enzyme domains as truly 'off-the-shelf' reagents to achieve a stereospecific synthesis. There is also the possibility of using this methodology for randomly combining DNA units that encode catalytic e.g. DH or transport e.g. ACP protein domains to generate combinatorial libraries of hybrid synthases. By using a suitable assay system to test for biological activity of the compounds that are generated by such means, it is

possible to go back and isolate the hybrid synthetic gene resposible for the production of these compounds.

From 6-methylsalicilic acid to maitotoxin, nature displays a staggering diversity in compounds that are synthesised by means of 'combinatorial gene-shuffling'. This methodology, or variations of this methodology can be used as effective tools towards harnessing the combinatorial potential of discrete enzymatic units or their sets that are the feature of multi-functional PKS and other systems.

A similar system to the *XbaI/dam* system described above, uses the restriction enzyme *FokI* which has the recognition site:

5'GGATG$(N)_9\downarrow$3'

3'CCTAC$(N)_{13}\uparrow$5'

with the *dcm* methylase of *E.coli.* Adding CCA or CCT to the 5' end of the *FokI* recognition site would make the site dcm sensitive. Furthermore, if the sequence TCTAGA were inserted into the redundant section of the *FokI* restriction site, then the enzyme could be used to generate '*XbaI*-cut ends'.

Methods

E. coli dam$^+$ DH10B^{TM} strain was purchased from Gibco BRL, USA.. *Pfu* DNA polymerase was purchased from Boeringer, Germany.

Construction of the final expression plasmid pAR10 was carried out in several steps, as follows. The ten PKS DNA units were amplified by PCR using *pfu* DNA polymerase. The respective regions of *eryAI* gene, as well as the oligonucleotides used for each PCR are outlined:

<u>LM</u> - segment of *eryAI* gene (Bevitt *et al.*, 1992) extending from nucleotide (N) 588 to N 2389;

5'GG**CATATG**GCGGACCTGTCAAAGCTCTCCGACAGT3' and

5'GG**TCTAGATC**CCAGCCGCGGTCGGTCGGCAGTCCCG3',

<u>KS1</u> - segment of *eryAI* gene extending from N 2384 to N 3769;

5'GG**TCTAGA**CTCGCTGTTCCACCCCGACCCCACGCGCTCGGGCACC GCGCACCA3' and

5'GG**TCTAGATC**GCGCAGCGCGGCGGACTCGTCGACGGGGGCGAAG GCGG3',

AT1 - segment of *eryAI* gene extending from N 3764 to N 4813;

5'GG**TCTAGA**CGGTCTCGCGACGGGAAACGCCGACGGTGCCGCCGTT GGAA3'

and

5'GG**TCTAGATC**CACCGCGACACCGGCGGCGAACGCGCGGGAGAGC GCTTCGC3',

KR1 - segment of *eryAI* gene extending from N 4808 to N 6316;

5'GG**TCTAGA**GTCGGTGCACCTGGGCACCGGAGCACGCCGGGTGCCC TT3'

and

5'GG**TCTAGATC**GTCGAAGAGCCTGGTCGGGCGCTGCGCGGTGTA3',

ACP1 - segment of *eryAI* gene extending from N 6311 to N 6679;

5'GG**TCTAGA**CGACGCGCGGCGGGCTGCGCCGCAGGCGCCGGCCGA ACCGCGGG3'

and

5'GG**TCTAGATC**GGCCGTGG-TCGCCGGTGCCGCCTGCTCGGCT3',

KS2 - segment of *eryAI* gene extending from N 6674 to N 8200;

5'GG**TCTAGA**CGAGCCGATCGCGATCGTCGGCATGGCGTGC- CGGCTGC3'

and

5'GG**TCTAGATC**GTGCACGGCCTCGGCGGTGTCGGCGGCGAGC- ACCGCGGCCCGCTCCTC3',

AT2 - segment of *eryAI* gene extending from N 8195 to N 9340;

5'GG**TCTAGA**GGCGGTGGCCGACGGCGCGGTGGTT3'

and

5'GG**TCTAGATC**GTCACGAGGGGTGGTGCGGTCCGGCAGCAGCCAGA A3',

KR2 - segment of *eryAI* gene extending from N 9335 to N 10639;

5'GG**TCTAGA**CGGCTGGTTCTACC-GGGTCGACTGGACCGAG3'

and

5'GG**TCTAGATC**CGGCCGGGGCCGGGCGGCGG-TGTAGGACT3',

ACP2 - segment of *eryAI* gene extending from N 10634 to N 10966; 5'GG**TCTAGA**CCGCATCGTCACGACCGCGCCGAGCGA3'

and

5'GG**TCTAGATC**G-GCGTCGAGGAAA3',

TE - segment of *eryAIII* gene (Donadio *et al.* 1991) extending from N 8753 to N 9602; 5'GG**TCTAGA**CAGCGGGACTCCCGCCCGGGAAGCG3'

and

5'GGGCTAGC**TCTAGATC**ATGAATTCCCTCCGCCCAGCCAGGCGTC3'.

All PCR products were 5' phosphorylated and ligated to *Sma*I-cut, dephosphorylated pUC18 vector and used to transform *E. coli* DH10B electrocompetent cells. The desired plasmids - containing the amplified DNA fragments were isolated and sequenced using standard pUC forward and reverse primers. No mistakes in the amplified products were detected. All ten plasmids were then used to transform the *E.coli* ET12567 dam$^-$ strain. Isolated DNA was digested with *Xba*I restriction enzyme and desired fragments isolated and purified. The TE unit was then ligated to *Xba*I-cut pCJR24 vector and the ligation products used to transform *E. coli* DH10B electrocompetent cells. Plasmid pAR1 was isolated, digested with *Xba*I, and ligated to the ACP2 fragment, and ligation products treated as mentioned above. The other DNA fragments, namely, KR2, AT2, KS2, ACP1, KR1, AT1 and KS1 were sequentially added to finally yield plasmid pAR10. This plasmid was then digested with *Nde*I and *Xba*I restriction enzymes and ligated with the LM fragment previously digested with the same two enzymes. The ligated products were used to transform *E. coli* DH10B electrocompetent cells and the final expression plasmid pAR10 isolated. Plasmid pAR10 was then used to transform *S. erythraea/JC2* strain and colonies carrying the expression plasmid were selected through resistance to thiostrepton upon integration of the plasmid into the S. *erythraea* chromosome. Single transformants were picked and grown on

tap-water medium plates supplemented with thiostrepton, following which single transformants were grown in 5X200ml of SM3 liquid media supplemented with 5 ug/ml of thiostrepton for seven days (Rowe *et al.*, 1998). Cells were removed by centrifugation, the supernatant was saturated with NaCl and extracted three times with equal volumes of ethyl acetate at pH 4.0. The solvent was evaporated to yield 1.12 g of crude product. A sample of this crude product was analysed by GC-MS. Two peaks were observed, corresponding to molecular masses 158 and 172, indicating the presence of the expected acetate- and propionate- derived polyketides (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-pentanoic acid δ-lactone (**1**) and (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-hexanoic acid δ-lactone (**2**). Compounds (1) and (2) were found to be structurally identical to those reported previously (Cortés *et al.*,1995).

*Characterisation of (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-pentanoic acid δ-lactone **(1)***

^{1}H NMR ($CDCl_3$, 500 MHz) $^{\delta}$H 4.45-4.35 (1H, dq, J = 6.56 and 1.62 Hz, C_5-H), 3.8 (1H, dd, J = 10.15 and 4.17 Hz C_3-H), 2.45-2.70 (1H, br, O-H), 2.42 (1H, dq, J = 10.0 and 6.97 Hz C_2-H), 2.05 (1H, m, C_4-H), 1.37 (3H, d, J = 7.17 Hz, C_2-CH_3), 1.32 (3H, d, J = 6.74 Hz, C_5-CH_3), 0.95 (3H, d, J = 7.20 Hz, C_4-CH_3) ppm. ^{13}C NMR ($CDCl_3$, 250 MHz) δ 174.20, 76.15, 73.62, 39.42, 38.14, 18.11, 14.24, 4.48.

*Characterisation of (2R,3S,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-hexanoic acid δ-lactone **(2)***

^{1}H NMR ($CDCl_3$, 500 MHz) dH 4.13 (1H, ddd, J = 8.12, 5.93 and 2.19 Hz, C_5-H), 3.82 (1H, m, C_3-H), 2.42-2.50 (1H, dq, J = 10.17 and 7.08 Hz, C_2-H), 2.12-2.19 (1H, m, C_4-H), 1.77-1.86 (1H, m, one of C_6-H_2), 1.52-1.61 (1H, m, one of C_6-H_2), 1.4 (3H, d, J = 7.09 Hz, C_2-CH_3), 1.0 (3H, t, J = 7.42 Hz, C_6-CH_3), 0.97 (3H, d, J = 6.96 Hz, C_4-CH_3) ppm. ^{13}C NMR ($CDCl_3$, 250 MHz) d 173.56, 81.34, 73.96, 40.08, 36.76, 25.27, 14.27, 9.88, 4.37.

<u>Example 2: *in vitro* assembly of DNA units</u>

Figure 7 outlines the strategy for the *in vitro* assembly of PKS DNA units. The inventors have constructed the multienzyme DEBS1-TE. The *in vivo* construction of the gene for DEBS1-TE, it should be noted, took 12 days to complete. The *in vitro* assembly on the other hand was completed in 2 days.

All ten domains, namely, LM, KS1, KR1, AT1, ACP1, KS2, AT2, KR2, ACP2 and TE were amplified by means of PCR. The forward primer in all cases, except the LM contained the *Spe*I recognition sequence 5'ACTAGT3' while the reverse primer was engineered in such a way that it contained the *Xba*I recognition sequence 5' TCTAGA3' and *Sma*I recognition sequence 5'CCCGGG3' downstream of the *Xba*I site (Figure 7). The amplification of the LM was carried out using a biotinylated forward primer and a reverse primer that contained the *Xba*I recognition sequence (5'TCTAGA3'). All the PCR products were cloned in pUC-18 vector and the resulting plasmids sequenced to detect possible errors introduced by PCR. All plasmids, except the one containing the LM unit were then digested with *Spe*I and *Sma*I, dephosphorylated in order to remove the 5' phosphate group and the appropriate fragments isolated and eluted. The LM unit was cleaved with *Xba*I and attached to a bead that was coated with streptavidin (following the manufacturer's instructions) as shown in figure 7.

The assembly process was initiated by adding DNA ligase to the tube containing a large excess of the first unit (KS1) and LM-bead. The reason for having a large excess of the KS1 unit compared to the LM-bead unit is to favour the LM-bead ligating to the incoming unit, as opposed to the self-ligation of the LM-bead (see figure 7). The ligation of the two DNA fragments is unidirectional as only the *Spe*I-cut end of KS1 complements the *Xba*I-cut end of the LM-bead. After the ligation was complete, the desired product of the ligation reaction, namely 'bead-LM-KS1' was separated from the reaction mixture and washed. This product was then cleaved with *Xba*I, in order to activate the 3' end of KS1. The beads were washed again to remove the small *Xba*I-*Sma*I DNA fragment that was

released from the 3' end of KS1 as a result of RE cleavage. The 'activated' bead-LM-KS1 unit was then ligated with *Spe*I, *Sma*I-cut and 5' dephosphorylated AT1. The *Spe*I-cut 5' end of AT1 complemented the *Xba*I-cut 3' end of KS1 to give bead-LM-KS1-AT1 as shown in figure 8. This product was separated from the reaction mixture and washed as before. The 3' end of AT1 in this product was then 'activated' through cleavage by *Xba*I, and the assembly process continued.

Finally, *Spe*I, *Sma*I-cut and 5' dephosphorylated TE unit was ligated with the DNA fragment that was now bead-LM-KS1-AT1-KR1-ACP1-KS2-AT2-KR2-ACP2 as shown in figure 9. The 3' end of the latter fragment was 'activated' by digesting it with *Xba*I. The assembled DEBS1-TE gene was then inserted in the expression plasmid pCJR24 and the resulting plasmid used to transform a streptomyces strain. The expected triketide lactone products were isolated and structurally characterised.

Use of the *in vitro* technology described above drastically reduces the time it takes to assemble predetermined or randomly shuffled genes. Also, the possibility of continuing with the assembly process while having numerous different assembly arrays attached to the beads, and splitting and mixing the beads between each unit/module addition from a library of units/modules, results finally in the generation of a cascade of different assemblies (Figure 10). These assembled genes can then be cloned simultaneously and expressed in a suitable host. An assay system can then be used to identify those assembled genes that yield bio-active compounds.

Example 3: Retrobiosynthetic synthesis of a target molecule

A strategy employing the invention in order to construct the highly potent anti-breast cancer drug discodermolide, the anticholesterol compound decarestrictine, and the antitumour compound octalacin using polyketide synthase domains/modules is outlined below.

<u>Discodermolide</u>

The drug discodermolide (Figure 18), isolated from the marine sponge '*Discodermia disoluta*', has been identified as a highly potent anti-cancer compound and 80 times more effective than the well known anticancer drug Taxol (TerHarr *et al.*, 1996). It has the same mechanism of action as Taxol, even though it is structurally different from the latter.

One can infer from its structure (Figure 18) that discodermolide is a polyketide and can therefore be constructed from a system that has the basic enzymatic building blocks (domains and modules) that make other polyketides like erythromycin and rapamycin. Having predicted that approximately 45 domains housed in 12 modules would be required in order to carry out the chemistry that accounts for the functionalities on the carbon skeleton of discodermolide, one can now begin to construct such a system. All one has to do is to identify the type and nature of the domains/modules that one requires to generate the observed functionalities, and then assemble these units in the desired order (Figure 18). The resulting DNA assembly can then be put into a bacterial strain that makes a functional polyketide synthase.

Until now, it would have been exceedingly difficult, if not impossible to assemble 45 or so pieces of DNA in the wanted order, for several reasons. Firstly, one would have to look for two different restriction enzymes every time one needed to assemble two DNA segments. This is because if one uses just one restriction enzyme at either end of the

domain, the already-assembled piece/pieces of DNA would be cleaved from the assembly every time one decided to insert a new domain. Secondly, in GC-rich DNA like the polyketide synthase producing *Streptomyces* strain, unique restriction enzyme sites are few and far between. To a molecular biologist, the task of assembling 40 pieces of DNA with the limitations mentioned above, would seem an insurmountable one. One would rather attempt to isolate the genes that make the drug at the first place than consider carrying out "step-by-step" reconstruction of the gene itself. In the case of discodermolide, even the last possibility is in the realms of fantasy. The organism within the marine sponge that makes the drug has not been identified. The only way discodermolide can be made available is through chemical synthesis - there have been a few chemical routes reported in literature recently (Marshall and Johns, 1998 and references therein). However, as is the case with most other complex molecules, large scale production of discodermolide, using the chemical route would turn out to be outrageously expensive. Chemists have been using the retrosynthetic analysis approach towards total synthesis of important bioactive molecules. This approach breaks the target compound into many smaller pieces - easily synthesised - which are then re-assembled.

The type of polyketide or other synthetic enzyme domains required in order to construct the target molecule from the starting units are identified using a "retrobiosynthetic analysis" approach for discodermolide,

by matching which molecules need to be condensed to form the macromolecule with the enzyme domains that carry out the required catalysis to build the macromolecule.

Having identified the enzyme units that are required, the unit-DNA segments are amplified using the polymerase-chain-reaction (PCR) - from the library of existing polyketide synthase unit-DNA, and the appropriate recognition sequences are attached to each unit-DNA fragment. All of the unit fragments are then replicated in a dam$^-$ strain whereby both the unmodified and modified sequences (5'TCTAGA3' and 5'**GA**TCTAGA3' respectively) are cleaved by the restriction enzyme *Xba*I.

Having constructed this library of appropriate PKS or other synthetic enzyme units, the corresponding DNA units are then assembled. The assembled DNA piece is then placed in a vector, so that it can be inserted in a bacterial strain to yield the desired synthetic protein. Suitable vectors have an antibiotic resistance marker (for selection of this vector on an antibiotic-rich media) and an "origin-of -replication" (ori). Ori is essential for the independent growth of the vector in any strain. Particularly suitable vectors for the expression of the synthetic enzymes of the invention are the actinomycete vectors described by Rowe *et al.* (1998).

The strain is then grown in a media that is supplemented with the antibiotic, the resistance gene for which is present in the vector.

Figures 4 and 5 show how the assembly proceeds. The first domain is inserted into a vector that is cut by cleavage with *Xba*I. After the ligation

of the domain has taken place with the vector, the DNA is put in a bacterial strain that is dam$^+$ and grown. Finally, bacterial colonies that have the desired vector-domain DNA are identified and DNA isolated from them. The whole procedure is cheap and fast. Only one restriction enzyme (*Xba*I) is made use of, routine cloning technology is employed, the desired DNA fragment is obtained, which can then be expressed in a Streptomyces strain to yield the polyketide synthase.

The *in vivo* "domain-by domain" construction of the discodermolide producing polyketide synthase would take approximately 55 days via this method. In comparison, assembly of modules would take less time, as one would need to assemble fewer pieces. Most importantly, once the synthase is shown to be functionally active, a large fermentation of the bacterial strain can be carried out, and the drug isolated in however much quantity one requires - unlike the chemical route where the starting materials have to be freshly synthesised every time one requires the target compound. Employing such a strategy would lead to a quick and inexpensive synthesis of important bioactive molecules like discodermolide.

Retrobiosynthetic analysis

The whole approach (retrobiosynthetic analysis followed by identification of PKS units, followed by assembly of PKS units) is made clearer in the following two examples.

Octalactin

A new addition to the rare class of eight-membered lactone natural products is the family of Octalactin. Octalactin A and B (Figure 20) are natural products isolated from the marine gorgonian octocoral '*Pacifigorgia sp.*' (Tapiolas et. al., 1991). Octalactin A shows very strong cytotoxicity toward B-16-F-10 murine melanoma and HCT-116 human colon tumour cell lines and is a promising drug candidate, while octalactine B displayed no such activity (Tapiolas et. al., 1991). Total syntheses of both octalactin A and B have been reported in literature. One such synthesis (Buszek, et. al., 1994) typically involves more than 12 chemical steps in leading to the target molecules. Clearly, large-scale production of octalactins using chemical synthesis is industrially not viable. On the other hand, the genes that code for the enzymes that make octalactins have not be identified or isolated. This means that at present, modified octalactins can only be made using chemical synthesis. A gene is constructed from the available PKS spare parts - that would code for the enzymes that would make octalactin B. Octalactin B can then be converted into the cytotoxic octalactin A by one-step stereospecific epoxidation. Also, once the gene for octalactin B is constructed and shown to make the octalactin PKS, genetic engineering on this gene would yield modified octalactin PKSs that in turn would synthesise octalactin analogues.

Clearly, a polyketide, the carbon skeleton of octalactin B (Figure 19) can be seen to be assembled by acetate and propionate units. The uptake

and assembly of these units in the prescribed sequence, as well as the functionalities that decorate the carbon chain of octalactin can be assigned to various PKS modules (see figure 19). Once a decision has been made regarding the type and nature of PKS modules, they can be strung together to make a gene using the invention. This gene can then be expressed in a suitable host in order to look for octalactin B production. The retrobiosynthetic approach towards octalactin is shown in detail in figure 19. A choice of what modules to select from the PKS module library is followed by amplification of the modular DNA fragments using the oligonucleotides such that the 5' and the 3' ends of every DNA fragment have the restriction enzyme recognition sites stated under the description of the invention. The choice of modules that, when assembled, would make the 'octalactin gene' is displayed as a schematic representation in figure 20.

Decarestrictine J

The molecule decarestrictine J can be synthesised using the retrebiosynthetic approach. Decarestrictine J is a ten-membered lactone that comes from the family of decarestrictines, shown to display strong anti-cholesterol activity (Grabley et. al., 1992). The total synthesis of Decarestrictine J has been reported and involves numerous chemical steps (Yamada et. al., 1995). The target molecule (figure 21) can be conceived to be formed by assembly of five acetate polyketide units. Using the retrobiosynthetic approach, one can identify the PKS domains/modules that

would be required for the carbon skeleton of decarestrictine J. A hypothetical decarestrcitine PKS is shown in figure 21. The loading module, as well as the four internal modules along with the TE domains can be conveniently assembled using the invention. The assembled 'decarestrictine gene' can then be expressed in a suitable host in order to check for the production of decarestrictine J.

In summary, the retrobiosynthetic approach involves the following steps;

a). Identification of the *number* and *nature* of carbon units that make up the target molecule

b). Identification of the modules/domains from libraries of polyketide/peptide synthetase/fatty acid/etc. encoding units that are responsible for the uptake of the said carbon units and the nature and degree of functionalisation of the carbon chain

c). Assembly of the said modules/domains using the methods of the invention

d). Expression of the assembled gene in a suitable expression host.

Example 4: Transforming strains with DNA encoding similar synthetic enzyme domains

A method for transforming expression strains with DNA encoding similar synthetic enzyme domains has been devised. Instead of using the TE PKS DNA fragment as a region of integration from the assembled gene into a streptomyces host (*S. erythraea*/JC2, Rowe *et al.*, 1998), a mutated *recA* gene fragment from streptomyces is used. The assembly process is carried

out in a *recA*⁻ *E. coli* strain (e.g. DH10B) as previously described. As this strain is recA⁻, one can assemble any number of identical DNA units. The vector, into which the assembled gene is being constructed, contains a portion of a streptomyces *recA* gene. This *recA* fragment carries a mutation. After the synthetic enzyme gene has been assembled, the vector is used to transform a streptomyces host (e.g. *S. lividans* or *S. erythraea*). The fragment of *recA* gene carrying a mutation recombines with the *recA* gene of the streptomyces host, abolishing the functional recA gene and making the strain recombination minus (Figure 11). This means that an event, such as the one described in figure 2 is now not possible. The strain is then grown to look for the encoded enzyme product. This strategy is tested by assembling a functional PKS gene having more than one type of identical DNA units (Figure 12).

Construction of the PKS multienzyme recDEBS1-TE

RecA protein has been characterised as a multifunctional enzyme that is essential for homologous recombination, DNA repair, SOS response and DNA rearrangements (Miller and Kokjohn, 1990). Most of the routinely used strains of *E. coli* are *recA*⁻. The gene for *recA* has been identified from many streptomyces strains. The first streptomyces *recA* gene to be characterised and isolated was from *S. lividans* (Nußbaumer and Wohlleben, 1994) *RecA* mutants have since been generated in *S. ambofaciens* (Aigle *et al.*, 1997). The streptomyces *recA* protein has approximately 372 amino acid residues (Figure 13). DNA sequence analysis suggests a coding region of 1122 bp, and is found to be highly conserved within streptomyces (Figure 14). In fact the *recA* mutants of *S. ambofaciens* were generated by integrating a mutated portion of the *S. lividans recA* gene into the *S. ambofaciens* host. It was found that a *recA* mutant lacking 30 aa from the C-terminus of the protein inhibited recombination events in *S. ambofaciens* (Aigle *et al.*, 1997).

A *recA* mutant of the streptomyces host that is used for expression of the assembled gene was generated.

The oligonucleotides:

5'- GG*TCTAGA*ATTCGGCAAGGGCGCCGGTCATGCGCAT-3' and

5'- GG*TCTAGATCT*GCGGCGTCGGCCGGGGCGGCGGAGGCG-3'

were used as the forward and reverse primers respectively and the 1000 bp internal region of *S. lividans recA* gene (Nußbaumer and Wohlleben, 1994) was amplified using *pfu* polymerase. An additional nucleotide (**C**) was incorporated into the forward primer to generate a frame shift in the amplified *recA* gene fragment. The PCR product was cloned in pUC-18 vector and sequenced to detect for possible errors during PCR. The 1.0 kbp *recA* fragment, flanked at both ends by an *Xba*I site was then inserted in the expression vector pCJR24 that has a unique *Xba*I site. The ligation mixture was used to transform *E. coli* DH10B cells and the desired plasmid DNA isolated. The resulting plasmid (pA*RecA*24) contains a non-methylated *Xba*I site at the 5' end of the *recA* gene fragment. The ten PKS DNA units, namely, TE, two each of ACP1, KR1, AT1 & KS1, and LM were inserted into the plasmid pA*RecA*24 to finally yield the expression plasmid p*RecA*D1TE. This plasmid was used to transform wild-type *S. lividans* protoplasts, and thiostrepton resistant colonies were grown in defined liquid media as described above. The compound (Figure 12) was isolated from the bacterial broth and chemically characterised.

Thus, it has been shown that a gene carrying interspaced DNA units that are identical in structure as well as function does not lead to internal recombination events, as the native *recA* gene of the streptomyces host has been disrupted. Furthermore, it has been shown that it is possible to use identical domains to reach the objective of generating hybrid synthetic enzyme systems. This strategy will greatly reduce the number of domains that otherwise have to be employed for the purposes of *de novo* PKS gene assembly that yields the desired chemical compounds. The inventors have established a set of 12 domains that are capable of functioning robustly and are independent of flexibility and spacial constraints - problems that beset the choice of domains and modules previously.

References

Aigle, B., Holl, A-C., Angulo, J.F., Leblond, P. and Decaris, B. (1997) Characterization of two *Streptomyces ambofaciens recA* mutants: identification of the *recA* protein by immunoblotting. *FEMS Microbiol. lett.*, **149**, 181-187.

Bevitt, D.J., Cortés, J., Haydock, S.F. and Leadlay, P.F. (1992) 6-Deoxyerythronolide B synthase 2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur. J. Biochem.*, **204**, 38-49.

Brown, M.J.B., Cortés, J., Cutter, A.L., Leadlay, P.F. and Staunton, J. (1995) A mutant generated by expression of an engineered DEBS1 protein from the erythromycin-producing polyketide synthase (PKS) in *Streptomyces coelicolor* produces the triketide as a lactone, but the major product is the nor-analogue derived from acetate as starter acid. *J. Chem. Soc., Chem. Commun.*, 1517-1518.

Buszek, KR., Sato, N. and Jeong, Y.M. (1994) Total synthesis of octalactin-A and octalactin-B. *J. Amer. Chem. Soc.* **116**, 5511-5512.

Carreras C. and Khosla C. (1998) Purification and *in vitro* reconstitution of the essential protein components of an aromatic polyketide synthase. *Biochemistry* **37**,2084-2088.

Cortés, J., Wiesmann, K.E.H., Roberts, G.A., Brown, M.J.B., Staunton, J. and Leadlay, P.F. (1995) Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science*, **268**, 1487-1489.

Donadio, S., McAlpine, J.B., Sheldon, P.J., Jackson, M. and Katz, L. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7119-7123.

Donadio, S., Staver, M.J., Mcalpine, J.B., Swanson, S.J. and Katz, L. (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science*, **252**, 675-679

Elsner, A., Engert, H., Saenger, W., Hamoen, L., Venema, G. and Bernhard, F. (1997) Substrate specificity of hybrid molecules from peptide synthetases. *J. Biol. Chem.* **272**, 4814-4819.

Fujimoto, D., Srinivasan, P.R. and Borek, E. (1965) On the nature of the deoxyribonucleic acid methylases. Biological evidence for the multiple nature of the enzymes. *Biochemistry* **4**, 2849-2855.

Geier, G. E. and Modrich, P. (1979) Recognition sequence of the *dam* methylase of *Escherichia coli* K12 and mode of cleavage of *Dpn* I endonuclease. *J. Biol. Chem,* **254**, 1408-1413.

Grabley, S., Granzer, E., Hutter, K., Ludwig, D., Mayer, M., Thiericke, R., Till, G., Wink, J., Phillips, S. and Zeeck, A. (1992) *J. Antibiot.* **45**, 56-65.

Jacobsen, J.R., Hutchinson, C.R., Cane, D.E. and Khosla, C. Precursor-directed biosynthesis of erythromycin analogs by an engineered polyketide synthase. *Science* **277**, 367-369 (1997)

Joshi, A.K. and Smith S. (1993) Construction of a cDNA encoding the multifunctional animal fatty acid synthase and expression in *Spodoptera frugiperda* cells using baculoviral vectors. *Biochem J.*,**296**, 143-149.

Kao, C.M., Luo, G.L., Katz, L., Cane, D.E. and Khosla, C. (1995) *J. Am. Chem. Soc.*, **117**, 9105-9106.

Kao, C.M., Luo, G.L., Katz, L., Cane, D.E. and Khosla, C. (1996) *J. Am. Chem. Soc.*, **118**, 9184-9185.

Kao, C.M., Luo, G.L., Katz, L., Cane, D.E. and Khosla, C., (1994) *J. Am. Chem. Soc.*, **116**, 11612-11613.

Kuhstoss, S., Huber, M., Turner, J.R., Paschal, J.W. and Rao, R.N. (1996) *Gene*, **183**, 231-236.

Lau, J., Fu, H., Cane, D. E. and Khosla, C. (1999) Dissecting the role of Acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry*, **38**, 1643-1651.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H. and MacNeil, T. (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesisutilizing a novel integration vector. *Gene* **111**, 61-68.

Marsden, A.F.A., Wilkinson, B., Cortés, J., Dunster, N.J., Staunton, J. and Leadlay, P.F. (1998) *Science*, **279**, 199-202.

Marshall, J.A. and Johns, B.A. (1998) Total synthesis of (+)-discodermolide. *J. Org. Chem.* **63**, 7885-7892.

McDaniel, R. et al. (1999) and references therein. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. USA*, **96**, 1846-1851.

Miller, R.V. and Kokjohn, T.A. (1990) General microbiology of *recA*: Environmental and evolutionary significance. *Annu. Rev. Microbiol.*, **44**, 365-394.

Nußbaumer, B. and Wohlleben, W. (1994) Identification, isolation and sequencing of the *recA* gene of *Streptomyces lividans* TK24. *FEMS Microbiol. lett.*, **118**, 51-56.

Oliynyk, M., Brown, M.J.B., Cortés, J., Staunton, J. and Leadlay, P.F. (1996) *Chem. Biol.*, **3**, 833-839.

Paitan, Y., Alon, G., Orr, E., Ron, E.Z., and Rosenberg, E. (1999) The first gene in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase. *J. Mol. Biol.* **286**,465-474.

Rowe, C.J., Cortés, J., Gaisser, S., Staunton, J., Leadlay, P.F. (1998) Construction of new vectors for regulated high-level expression in actinomycetes. *Gene*, **216**, 215-223.

Ruan, X., Pereda, A., Stassi, D.L., Zeidner, D., Summers, R.G., Jackson, M., Shivakumar, A., Kakavas, S., Staver, M.J., Donadio, S. and Katz, L. (1997) Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *J. Bacteriol.* **179**, 6416-6425.

Shen, B., Du, L., Sanchez, C., Chen, M. and Edwards, D.J. (1999) Bleomycin biosynthesis in Streptomyces verticillus ATCC15003: A model of hybrid peptide and polyketide biosynthesis. *Bioorganic Chemistry* **27**, 123-129.

Tapiolas, D.M., Roman, M., Fenical, W., Stout, T.J. and Clardy, J. (1991)

Octalactin-A and Octalactin-B - cytotoxic 8-membered-ring lactones from a marine bacterium, *Streptomyces sp. J. Amer. Chem. Soc.* **113**, 4682-4683.

TerHaar, E., Kowalski, R.J., Hamel, E., Lin, C.M., Longley, R.E., Gunasekera, S.P., Rosenkranz, H.S. and Day, B.W. (1996) Discodermolide, a cytotoxic marine agent that stabilizes microtubules more potently than taxol. *Biochemistry* **35**, 243-250.

Yamada, S., Tanaka, A. and Oritani, T. (1995) Total synthesis of Decarestrictine-J. *Biosci. Biotech. & Biochem.* **59**, 1657-1660

Ziermann, R. and Betlach, M.C. (1999) Recombinant Polyketide Synthesis in *Streptomyces*: Engineering of improved host strains. *BioTechniques* **26**, 106-110.

CLAIMS

1. A method of assembling several DNA units in sequence in a DNA construct, which method comprises the steps of

a) providing each DNA unit with a restriction enzyme recognition sequence at it's 5' end and with a recognition sequence for the same restriction enzyme at its 3' end that is combined with a recognition site for a DNA modification enzyme.

b) providing a starting DNA construct having an accessible restriction site for the same or a compatible restriction enzyme and cleaving the starting DNA construct with such a restriction enzyme,

c) inserting the desired DNA unit and bringing the ligated product into contact with a DNA modification enzyme such that the restriction site at the 3' end of the inserted DNA unit is abolished

d) cleaving the ligated product at an accessible unmodified recognition site for the same or a compatible restriction enzyme,

e) repeating steps c) and d) to introduce each desired DNA unit to give a DNA construct containing all the desired units in sequence.

2. The method of claim 1 wherein the DNA modification enzyme is a methylase.

3. The method of claim 2 wherein the methylase is the *dam* methylase of *Escherichia coli.*

4. The method of claim 3 which comprises the steps of

a) providing each DNA unit with an *Xba*I recognition sequence 5'XXTCTAGA3' (where XX is not GA) at it's 5' end and with an *Xba*I recognition sequence 5'GATCTAGA3' at its 3' end.

b) providing a starting DNA construct having an accessible *Xba*I site and cleaving the starting DNA construct with *Xba*I,

c) inserting the desired DNA unit and using a resulting ligated product to transform a dam+ strain of *E. coli*,

d) recovering a resulting plasmid and cleaving the plasmid at an accessible *Xba*I site with *Xba*I,

e) repeating steps c) and d) to introduce each desired DNA unit to give a DNA construct containing all the desired units in sequence.

5. The method of any one of claims 1 to 4, wherein the recognition sequences for the restriction enzyme and the DNA modification enzyme are created in the DNA units prior to cutting with the restriction enzyme.

6. The method of claim 5 wherein the restriction sites are created in the fragment by means of a primer extension reaction.

7. The method of any one of claims 1 to 6, wherein the DNA construct is an expression vector capable of facilitating expression of the protein encoded by the desired DNA units

8. The method of claim 3 or claim 4, wherein the DNA modification is removed and the restriction site re-established by replicating the ligated product in a dam- strain of *E. coli* by means of a suitable vector.

9. A method of making an assembly of several DNA units in sequence which method comprises the steps of:

a) providing a first DNA unit with a recognition sequence for a first restriction enzyme at its 3' end, and cleaving the said first DNA unit with said first restriction enzyme,

b) providing each other DNA unit with a recognition sequence at its 5' end for a second restriction enzyme which has a compatible ligation sequence with that of the first restriction enzyme, and a downstream recognition sequence for said first restriction enzyme followed by a downstream recognition sequence for a third restriction enzyme at its 3' end, and cleaving each said other DNA unit with the second and third restriction enzymes,

c) ligating the said first DNA unit with a desired other DNA unit to form a ligated product such that the ligation of the two units abolishes the recognition site for the first restriction enzyme at the ligation junction, and cleaving the ligated product with said first restriction enzyme,

d) ligating the product from c) with a desired DNA unit from b) to form a ligated product and cleaving the ligated product with said first restriction enzyme

e) repeating step d) with each other DNA unit in tum so as to assemble the DNA units in sequence.

10. The method of claim 9 which method comprises the steps of:

a) providing a first DNA unit with an *Xba*I recognition sequence 5'TCTAGA3' at its 3' end, and cleaving the said first DNA unit with *Xba*I,

b) providing each other DNA unit with a *Spe*I recognition sequence 5'ACTAGT3' at its 5' end, and a downstream *Xba*I recognition sequence 5'TCTAGA3' followed by a downstream *Sma*I recognition sequence 5'CCCGGG3' at its 3' end, cleaving each said other DNA unit with *Spe*I and *Sma*I, and dephosphorylating the 5' end of the cleaved DNA unit,

c) ligating the said first DNA unit with a desired other DNA unit to form a ligated product and cleaving the ligated product with *Xba*I,

d) ligating the product from c) with a desired DNA unit from b) to form a ligated product and cleaving the ligated product with *Xba*I

e) repeating step d) with each other DNA unit in turn so as to assemble the DNA units in sequence.

11. The method of claim 9 or claim 10 wherein the assembly occurs via stepwise addition of fragments to a vector

12. The method of claim 9 or claim 10 wherein the said first DNA unit is attached to the solid phase for use in step c)

13. The method of claim 12, wherein the solid phase is split and mixed between steps c), d), and e) to make several different assemblies.

14. The method of any one of claims 9-13, wherein the recognition sequences in one or more of the DNA units are introduced by means of extension primers.

15. The method of any one of claims 9-14 wherein the assembly of several DNA units is inserted in to an expression vector which is used to transform a host capable of expressing the protein encoded by the vector

16. The method of any one of claims 1-15, wherein one or more of the DNA units encodes a catalytic or transport protein domain. *(see Kleinkauf peptide/polyketide systems paper)*

17. The method of claim 16 wherein one or more of the DNA units are derived from polyketide synthesising enzyme domain DNA sequences.

18. The method of claim 16 wherein one or more of the DNA units are derived from peptide synthesising enzyme domain DNA sequences.

19. The method of claim 16 wherein one or more of the DNA units are derived from hybrid peptide polyketide enzyme domain DNA sequences.

20. The method of claim 16 wherein one or more of the DNA units are derived from fatty acid synthesising enzyme domain DNA sequences

21. The method of claim 16 wherein one or more of the DNA units encode modules comprising one or more catalytic or transport domains

22. DNA constructs incorporating one or more DNA assemblies encoding synthetic enzymes made by any one of the methods of claims 1-21.

23. Synthetic enzymes encoded by one or more DNA assemblies made by the methods of anyone of claims 1-21

24. Hosts expressing DNA constructs encoding one or more synthetic enzymes made by any one of the methods of claims 1-21.

25. Hybrids of transformed hosts expressing one or more DNA constructs encoding synthetic enzymes incorporating a DNA assembly made by any one of the methods of claims 1-21.

26. Compounds produced by synthetic enzymes encoded by DNA assemblies made by any one of the methods of claims 1-21.

27. A method of synthesising a target molecule comprising the steps of

a) examining the composition and stereochemistry of a target molecule,

b) determining which catalytic and transport domains need to be present in a synthetic enzyme in order to catalyse the synthesis of the target molecule,

c) using any one of the methods of claims 1-21 to assemble the required DNA units encoding the catalytic and transport domains into a

DNA assembly that encodes said synthetic enzyme which is capable of synthesising the target molecule.

d) placing the DNA assembly into a vector to allow expression of the synthetic enzyme in a host capable of synthesising the target molecule after transformation with said vector.

28. The method of claim 27 wherein the transformed host is tested for the presence of the target molecule after step d).

29. The transformed host of claim 27.

30. Use of transformed host of claim 27 to produce said target molecule.

31. A method of making a synthetic enzyme to catalyse the synthesis of a target molecule comprising the steps of

a) examining the composition and stereochemistry of a target molecule,

b) determining which catalytic and transport domains need to be present in the synthetic enzyme in order to catalyse the synthesis of the target molecule,

c) using any one of the methods of claims 1-21 to assemble the required DNA units encoding the catalytic and transport domains into a DNA assembly that encodes an enzyme which is capable of synthesising the target molecule.

d) expressing the DNA assembly in a suitable host to produce the enzyme.

32. A library of DNA units encoding catalytic or transport protein domains, wherein each DNA unit has a recognition sequence for a restriction enzyme at it's 5'-end and a second recognition sequence for the same or a compatible enzyme at it's 3'-end which incorporates a recognition sequence for a DNA modifying enzyme.

33. The library of claim 32, wherein each DNA unit has an *Xba*I recognition sequence 5'XXTCTAGA3' (where XX is not GA) at it's 5'-end and an *Xba*I recognition sequence 5'GATCTAGA3' at it's 3'-end

34. A library of DNA units encoding catalytic or transport protein domains, wherein each DNA unit has a recognition sequence at its 5' end for a first restriction enzyme, and a downstream recognition sequence for a second restriction enzyme followed by a downstream recognition sequence for a third restriction enzyme at its 3' end, such that the DNA units, once restricted by the first and second restriction enzymes can be ligated together to abolish the restriction sites at the ligation junction.

35. The library of claim 34, wherein each DNA unit has a *Spe*I recognition sequence 5'ACTAGT3' at its 5'-end, and a downstream *Xba*I recognition sequence 5'TCTAGA3' followed by a downstream *Sma*I recognition sequence 5'CCCGGG3' at it's 3'-end

34. The library of claim 32 or claim 34, wherein the DNA units encode polyketide synthetic domains, comprising two KS domains, at least two AT domains, two KR domains, two DH domains, two ER domains, an ACP domain and a TE domain.

35. A module comprising a DNA sequence encoding a functional set of polyketide synthetic domains wherein the module has a recognition sequence for a restriction enzyme at it's 5'-end and a second recognition sequence for the same or a compatible enzyme at it's 3'-end which incorporates a recognition sequence for a DNA modifying enzyme

36. The module as claimed in claim 35, wherein the module has an *Xba*I recognition sequence 5'XXTCTAGA3' (where XX is not GA) at it's 5'-end and an *Xba*I recognition sequence 5'GATCTAGA3' at it's 3'-end

37. A module comprising a DNA sequence encoding a functional set of polyketide synthetic domains wherein the module has a recognition sequence at its 5' end for a first restriction enzyme, and a downstream recognition sequence for a second restriction enzyme followed a downstream recognition sequence for a third restriction enzyme at its 3' end, such that the DNA units, once restricted by the first and second restriction enzymes can be ligated together to abolish the restriction sites at the ligation junction

38. The module as claimed in claim 37, wherein the module has a *Spe*I recognition sequence 5'ACTAGT3' at its 5'-end, and a downstream *Xba*I recognition sequence 5'TCTAGA3' followed by a downstream *Sma*I recognition sequence 5'CCCGGG3' at it's 3'-end

39. A module as claimed in claim 35 or claim 37, wherein the DNA units encode polyketide synthetic domains, comprising two KS domains, at least two AT domains, two KR domains, two DH domains, two ER domains, an ACP domain and a TE domain

40. A vector containing one or more modules as claimed in claim 35 or claim 37.

41. The vector as claimed in claim 40, wherein a non-functional recA gene is also present.

42. A method of transforming a host with one or more synthetic DNA assemblies encoding enzyme domains which comprises the steps of:

a) Inserting said DNA assembly into a vector containing a mutated internal fragment of a recA gene sequence such that the vector is capable of undergoing homologous recombination with the recA gene of the host,

b) bringing said vector into contact with a host chromosome under conditions which permit homologous recombination to take place,

c) disrupting the host recA gene by the integration of the DNA of said vector into the chromosome.

43. The method of claim 42 wherein the expression vector is used to transform a Steptomyces host.

44. The method of claim 42 or claim 43, wherein the DNA assemblies are modules according to claim 35 or claim 37.

45. A host lacking a recA function, transformed with a vector containing one or more modules according to claim 35 or 37.

46. A kit containing DNA units, DNA modules, vectors, DNA manipulation hosts, DNA modification hosts, expression hosts, or solid phase elements for use in the methods claimed herein.